

Supplementary Information

Speech Foundation Models for Parkinson’s Disease Detection: A Layer-Wise Comparison with Handcrafted Acoustics Across Two Cohorts

Hadi Sedigh Malekroodi¹, Byeong-il Lee^{1,2,3,*}, and Myunggi Yi^{1,2,4,*}

¹Industry 4.0 Convergence Bionics Engineering, Pukyong National University, Busan 48513, Republic of Korea

²Digital Healthcare Research Center, Institute of Information Technology and Convergence, Pukyong National University, Busan 48513, Republic of Korea

³Major of Human Bioconvergence, Division of Smart Healthcare, Pukyong National University, Busan 48513, Republic of Korea

⁴Major of Biomedical Engineering, Division of Smart Healthcare, Pukyong National University, Busan 48513, Republic of Korea

*Corresponding authors: bilee@pknu.ac.kr; myunggi@pknu.ac.kr

This Supplementary Information provides material referenced but not fully detailed in the paper: (1) segment- and subject-level classification results for all four downstream classifiers per model, extending the main-text tables, which report only the best classifier per model; (2) subject-level classification results for the same model–classifier pairs stratified by gender and age stratum; (3) the full PC-GITA and NeuroVoz reading sentences, the eGeMAPS feature inventory, and the downstream classifier hyperparameter settings; and (4) additional segment-level performance figures, including the model-wise *F1*-score comparison across representations, per-sentence ASR decoding errors, accuracy and ROC-AUC distributions, layer-wise ROC-AUC trajectories, and sustained-vowel demographic parity. Table and figure numbers without an “S” prefix refer to the main manuscript.

Supplementary Tables

Table S1: Full binary classification performance (all four downstream classifiers per model) on Sentence Reading tasks for the PC-GITA and NeuroVoz cohorts, at each model–classifier pair’s best-performing layer (selected by segment-level *F1*-score). This extends Table 1 in the main text, which reports only the best classifier per model. Values in parentheses are the bootstrap margin of error at a 95% confidence level.

Model	Classifier	Best Layer	Segment-level			Subject-level		
			Acc	F1	AUC	Acc	F1	AUC
PC-GITA Sentences (Colombian Spanish)								
Whisper Small	SVM	6	0.8200 (0.0225)	0.8196 (0.0265)	0.8824 (0.0205)	0.8200 (0.0734)	0.8163 (0.0789)	0.9064 (0.0589)
Whisper Small	XGB	6	0.7990 (0.0243)	0.8035 (0.0261)	0.8736 (0.0206)	0.8500 (0.0702)	0.8571 (0.0721)	0.9120 (0.0612)
Whisper Small	RF	4	0.8120 (0.0237)	0.8157 (0.0245)	0.8934 (0.0195)	0.8500 (0.0683)	0.8515 (0.0744)	0.9304 (0.0534)

(Supplementary Table S1 continued)

Model	Classifier	Best Layer	Segment-level			Subject-level		
			Acc	F1	AUC	Acc	F1	AUC
Whisper Small	LR	2	0.8300 (0.0241)	0.8317 (0.0252)	0.8850 (0.0218)	0.8400 (0.0665)	0.8431 (0.0709)	0.9096 (0.0588)
Whisper Medium	SVM	7	0.8380 (0.0226)	0.8383 (0.0244)	0.9073 (0.0188)	0.8500 (0.0694)	0.8544 (0.0748)	0.9260 (0.0514)
Whisper Medium	XGB	6	0.8170 (0.0236)	0.8201 (0.0254)	0.8877 (0.0198)	0.8700 (0.0647)	0.8738 (0.0667)	0.9244 (0.0541)
Whisper Medium	RF	7	0.8330 (0.0228)	0.8322 (0.0248)	0.8994 (0.0188)	0.8600 (0.0670)	0.8600 (0.0741)	0.9332 (0.0483)
Whisper Medium	LR	6	0.8320 (0.0226)	0.8327 (0.0246)	0.9010 (0.0198)	0.8500 (0.0668)	0.8515 (0.0718)	0.9216 (0.0532)
Whisper Large	SVM	14	0.8340 (0.0234)	0.8323 (0.0255)	0.9195 (0.0170)	0.8400 (0.0704)	0.8431 (0.0753)	0.9360 (0.0464)
Whisper Large	XGB	12	0.8280 (0.0227)	0.8205 (0.0264)	0.8919 (0.0204)	0.8900 (0.0612)	0.8842 (0.0691)	0.9244 (0.0552)
Whisper Large	RF	13	0.8330 (0.0229)	0.8338 (0.0242)	0.8999 (0.0191)	0.8600 (0.0698)	0.8627 (0.0715)	0.9340 (0.0482)
Whisper Large	LR	14	0.8100 (0.0241)	0.8123 (0.0263)	0.8928 (0.0191)	0.8400 (0.0682)	0.8431 (0.0728)	0.9168 (0.0507)
Qwen3-ASR 0.6B	SVM	8	0.8660 (0.0209)	0.8649 (0.0225)	0.9214 (0.0179)	0.9000 (0.0572)	0.8980 (0.0633)	0.9328 (0.0522)
Qwen3-ASR 0.6B	XGB	3	0.8210 (0.0254)	0.8250 (0.0255)	0.8793 (0.0226)	0.8400 (0.0724)	0.8462 (0.0752)	0.9056 (0.0638)
Qwen3-ASR 0.6B	RF	3	0.8270 (0.0231)	0.8292 (0.0243)	0.8894 (0.0221)	0.8600 (0.0704)	0.8600 (0.0748)	0.9148 (0.0640)
Qwen3-ASR 0.6B	LR	4	0.8250 (0.0233)	0.8289 (0.0250)	0.8924 (0.0192)	0.8400 (0.0672)	0.8462 (0.0684)	0.9120 (0.0562)
Qwen3-ASR 1.7B	SVM	9	0.8810 (0.0196)	0.8787 (0.0212)	0.9229 (0.0180)	0.9000 (0.0571)	0.8980 (0.0628)	0.9352 (0.0501)
Qwen3-ASR 1.7B	XGB	7	0.8130 (0.0237)	0.8074 (0.0265)	0.8820 (0.0206)	0.8600 (0.0702)	0.8542 (0.0795)	0.9184 (0.0568)
Qwen3-ASR 1.7B	RF	6	0.8190 (0.0237)	0.8151 (0.0265)	0.8804 (0.0213)	0.8600 (0.0644)	0.8600 (0.0709)	0.9072 (0.0616)
Qwen3-ASR 1.7B	LR	9	0.8460 (0.0224)	0.8451 (0.0242)	0.9048 (0.0194)	0.8700 (0.0667)	0.8687 (0.0725)	0.9196 (0.0572)
Nemotron 3.5 0.6B	SVM	5	0.8430 (0.0210)	0.8419 (0.0233)	0.9040 (0.0196)	0.8500 (0.0670)	0.8485 (0.0736)	0.9152 (0.0577)
Nemotron 3.5 0.6B	XGB	10	0.7950 (0.0231)	0.8000 (0.0254)	0.8719 (0.0223)	0.8400 (0.0689)	0.8431 (0.0768)	0.9000 (0.0610)
Nemotron 3.5 0.6B	RF	4	0.8090 (0.0237)	0.8092 (0.0273)	0.8766 (0.0212)	0.8400 (0.0704)	0.8431 (0.0745)	0.9100 (0.0593)
Nemotron 3.5 0.6B	LR	4	0.8250 (0.0230)	0.8269 (0.0254)	0.8809 (0.0210)	0.8500 (0.0691)	0.8515 (0.0737)	0.9032 (0.0650)
Wav2Vec 2.0 (XLS-R)	SVM	21	0.8260 (0.0231)	0.8249 (0.0253)	0.8837 (0.0209)	0.8700 (0.0653)	0.8713 (0.0667)	0.8872 (0.0679)
Wav2Vec 2.0 (XLS-R)	XGB	5	0.7900 (0.0255)	0.7862 (0.0288)	0.8573 (0.0241)	0.8500 (0.0680)	0.8485 (0.0748)	0.9008 (0.0619)
Wav2Vec 2.0 (XLS-R)	RF	5	0.7950 (0.0250)	0.7867 (0.0287)	0.8625 (0.0239)	0.8500 (0.0683)	0.8454 (0.0757)	0.9056 (0.0613)
Wav2Vec 2.0 (XLS-R)	LR	3	0.7770 (0.0245)	0.7768 (0.0273)	0.8492 (0.0230)	0.8200 (0.0734)	0.8200 (0.0820)	0.8692 (0.0716)
OmniASR-S	SVM	9	0.8140 (0.0232)	0.8147 (0.0240)	0.8695 (0.0211)	0.8400 (0.0726)	0.8367 (0.0762)	0.8972 (0.0614)
OmniASR-S	XGB	17	0.7760 (0.0253)	0.7742 (0.0277)	0.8635 (0.0217)	0.8100 (0.0753)	0.8155 (0.0787)	0.9120 (0.0518)
OmniASR-S	RF	17	0.7840 (0.0244)	0.7866 (0.0269)	0.8560 (0.0218)	0.7900 (0.0791)	0.7921 (0.0841)	0.9172 (0.0488)
OmniASR-S	LR	16	0.7920 (0.0242)	0.7953 (0.0267)	0.8359 (0.0254)	0.8200 (0.0738)	0.8200 (0.0800)	0.8752 (0.0686)
OmniASR-M	SVM	9	0.8060 (0.0243)	0.8072 (0.0263)	0.8762 (0.0221)	0.8300 (0.0737)	0.8283 (0.0822)	0.9032 (0.0588)
OmniASR-M	XGB	33	0.7840 (0.0249)	0.7886 (0.0263)	0.8681 (0.0209)	0.8100 (0.0760)	0.8190 (0.0808)	0.9084 (0.0528)
OmniASR-M	RF	32	0.7860 (0.0245)	0.7860 (0.0265)	0.8533 (0.0232)	0.8500 (0.0699)	0.8515 (0.0748)	0.9032 (0.0634)
OmniASR-M	LR	10	0.7990 (0.0249)	0.8000 (0.0266)	0.8596 (0.0237)	0.8500 (0.0696)	0.8485 (0.0727)	0.9048 (0.0639)
eGeMAPS (openSMILE)	SVM	0	0.7420 (0.0266)	0.7461 (0.0295)	0.8004 (0.0264)	0.7700 (0.0830)	0.7723 (0.0889)	0.8312 (0.0838)
eGeMAPS (openSMILE)	XGB	0	0.7110 (0.0271)	0.7197 (0.0303)	0.7762 (0.0283)	0.7700 (0.0812)	0.7810 (0.0845)	0.8400 (0.0812)
eGeMAPS (openSMILE)	RF	0	0.7200 (0.0270)	0.7323 (0.0287)	0.7896 (0.0270)	0.7600 (0.0787)	0.7818 (0.0845)	0.8456 (0.0800)

(Supplementary Table S1 continued)

Model	Classifier	Best Layer	Segment-level			Subject-level		
			Acc	F1	AUC	Acc	F1	AUC
eGeMAPS (openSMILE)	LR	0	0.7320 (0.0265)	0.7347 (0.0300)	0.7823 (0.0287)	0.8000 (0.0762)	0.8000 (0.0844)	0.8400 (0.0821)
NeuroVoz Sentences (Castilian Spanish)								
Whisper Small	SVM	6	0.8433 (0.0171)	0.8372 (0.0188)	0.9265 (0.0120)	0.8872 (0.0563)	0.8824 (0.0642)	0.9654 (0.0291)
Whisper Small	XGB	4	0.7997 (0.0187)	0.7934 (0.0212)	0.8737 (0.0167)	0.8685 (0.0594)	0.8627 (0.0695)	0.9227 (0.0511)
Whisper Small	RF	11	0.7991 (0.0189)	0.7929 (0.0207)	0.8791 (0.0152)	0.9059 (0.0554)	0.9020 (0.0606)	0.9584 (0.0346)
Whisper Small	LR	12	0.8482 (0.0178)	0.8431 (0.0192)	0.9280 (0.0112)	0.9439 (0.0430)	0.9423 (0.0471)	0.9794 (0.0253)
Whisper Medium	SVM	12	0.8505 (0.0157)	0.8454 (0.0179)	0.9320 (0.0107)	0.8963 (0.0570)	0.8911 (0.0637)	0.9682 (0.0280)
Whisper Medium	XGB	9	0.8213 (0.0180)	0.8136 (0.0205)	0.8907 (0.0152)	0.8776 (0.0600)	0.8713 (0.0681)	0.9479 (0.0375)
Whisper Medium	RF	18	0.8067 (0.0183)	0.8050 (0.0204)	0.8770 (0.0158)	0.8787 (0.0613)	0.8762 (0.0665)	0.9503 (0.0363)
Whisper Medium	LR	8	0.8563 (0.0152)	0.8506 (0.0176)	0.9332 (0.0107)	0.8969 (0.0601)	0.8932 (0.0648)	0.9675 (0.0317)
Whisper Large	SVM	19	0.8576 (0.0165)	0.8528 (0.0180)	0.9425 (0.0097)	0.9156 (0.0513)	0.9126 (0.0559)	0.9755 (0.0222)
Whisper Large	XGB	19	0.8158 (0.0175)	0.8116 (0.0210)	0.8949 (0.0146)	0.8781 (0.0601)	0.8738 (0.0667)	0.9490 (0.0378)
Whisper Large	RF	19	0.8083 (0.0186)	0.8053 (0.0203)	0.8856 (0.0161)	0.8787 (0.0612)	0.8762 (0.0680)	0.9434 (0.0399)
Whisper Large	LR	20	0.8526 (0.0169)	0.8499 (0.0183)	0.9343 (0.0110)	0.8974 (0.0583)	0.8952 (0.0628)	0.9710 (0.0264)
Qwen3-ASR 0.6B	SVM	10	0.8770 (0.0148)	0.8721 (0.0165)	0.9542 (0.0084)	0.9348 (0.0430)	0.9333 (0.0492)	0.9811 (0.0217)
Qwen3-ASR 0.6B	XGB	10	0.8306 (0.0184)	0.8265 (0.0215)	0.9099 (0.0133)	0.8781 (0.0599)	0.8738 (0.0699)	0.9647 (0.0286)
Qwen3-ASR 0.6B	RF	10	0.8218 (0.0179)	0.8177 (0.0199)	0.9006 (0.0142)	0.9252 (0.0482)	0.9231 (0.0526)	0.9706 (0.0272)
Qwen3-ASR 0.6B	LR	10	0.8743 (0.0162)	0.8704 (0.0181)	0.9463 (0.0098)	0.9252 (0.0485)	0.9231 (0.0518)	0.9829 (0.0179)
Qwen3-ASR 1.7B	SVM	13	0.8818 (0.0152)	0.8774 (0.0162)	0.9579 (0.0078)	0.9241 (0.0492)	0.9200 (0.0550)	0.9804 (0.0227)
Qwen3-ASR 1.7B	XGB	13	0.8526 (0.0166)	0.8452 (0.0187)	0.9192 (0.0132)	0.9049 (0.0527)	0.8980 (0.0620)	0.9790 (0.0239)
Qwen3-ASR 1.7B	RF	14	0.8434 (0.0162)	0.8376 (0.0187)	0.9157 (0.0127)	0.9054 (0.0534)	0.9000 (0.0616)	0.9755 (0.0234)
Qwen3-ASR 1.7B	LR	23	0.8803 (0.0157)	0.8770 (0.0173)	0.9498 (0.0092)	0.9156 (0.0518)	0.9126 (0.0581)	0.9794 (0.0203)
Nemotron 3.5 0.6B	SVM	10	0.8744 (0.0152)	0.8712 (0.0164)	0.9428 (0.0102)	0.9348 (0.0437)	0.9333 (0.0476)	0.9762 (0.0260)
Nemotron 3.5 0.6B	XGB	8	0.8278 (0.0179)	0.8203 (0.0195)	0.9037 (0.0139)	0.8675 (0.0618)	0.8571 (0.0730)	0.9570 (0.0330)
Nemotron 3.5 0.6B	RF	10	0.8180 (0.0183)	0.8125 (0.0202)	0.8937 (0.0146)	0.8963 (0.0559)	0.8911 (0.0626)	0.9587 (0.0353)
Nemotron 3.5 0.6B	LR	6	0.8633 (0.0158)	0.8570 (0.0178)	0.9386 (0.0098)	0.9241 (0.0502)	0.9200 (0.0566)	0.9787 (0.0197)
Wav2Vec 2.0 (XLS-R)	SVM	12	0.8856 (0.0149)	0.8828 (0.0160)	0.9543 (0.0088)	0.9247 (0.0482)	0.9216 (0.0522)	0.9769 (0.0239)
Wav2Vec 2.0 (XLS-R)	XGB	12	0.8430 (0.0171)	0.8346 (0.0197)	0.9067 (0.0135)	0.8963 (0.0563)	0.8911 (0.0635)	0.9643 (0.0282)
Wav2Vec 2.0 (XLS-R)	RF	13	0.8362 (0.0168)	0.8300 (0.0190)	0.9105 (0.0133)	0.9156 (0.0514)	0.9126 (0.0580)	0.9650 (0.0292)
Wav2Vec 2.0 (XLS-R)	LR	11	0.8788 (0.0149)	0.8770 (0.0161)	0.9485 (0.0093)	0.9161 (0.0515)	0.9143 (0.0541)	0.9790 (0.0236)
OmniASR-S	SVM	10	0.8346 (0.0182)	0.8291 (0.0196)	0.9076 (0.0136)	0.9059 (0.0528)	0.9020 (0.0599)	0.9626 (0.0279)
OmniASR-S	XGB	9	0.7804 (0.0193)	0.7702 (0.0232)	0.8557 (0.0176)	0.8402 (0.0682)	0.8317 (0.0750)	0.9455 (0.0379)
OmniASR-S	RF	17	0.7729 (0.0188)	0.7701 (0.0211)	0.8524 (0.0167)	0.8878 (0.0575)	0.8846 (0.0621)	0.9392 (0.0439)
OmniASR-S	LR	15	0.8176 (0.0182)	0.8129 (0.0204)	0.8908 (0.0148)	0.9065 (0.0519)	0.9038 (0.0597)	0.9647 (0.0335)
OmniASR-M	SVM	29	0.8251 (0.0178)	0.8196 (0.0196)	0.9061 (0.0131)	0.8589 (0.0616)	0.8515 (0.0724)	0.9559 (0.0338)
OmniASR-M	XGB	19	0.7821 (0.0197)	0.7711 (0.0224)	0.8686 (0.0160)	0.8958 (0.0554)	0.8889 (0.0631)	0.9790 (0.0189)

(Supplementary Table S1 continued)

Model	Classifier	Best Layer	Segment-level			Subject-level		
			Acc	F1	AUC	Acc	F1	AUC
OmniASR-M	RF	30	0.7890 (0.0194)	0.7876 (0.0223)	0.8670 (0.0175)	0.8883 (0.0579)	0.8868 (0.0653)	0.9528 (0.0384)
OmniASR-M	LR	10	0.8096 (0.0195)	0.8074 (0.0215)	0.8902 (0.0153)	0.9257 (0.0502)	0.9245 (0.0512)	0.9629 (0.0339)
eGeMAPS (openSMILE)	SVM	0	0.7314 (0.0214)	0.7202 (0.0254)	0.8002 (0.0215)	0.7937 (0.0768)	0.7843 (0.0917)	0.8650 (0.0700)
eGeMAPS (openSMILE)	XGB	0	0.7204 (0.0218)	0.7032 (0.0269)	0.7916 (0.0218)	0.8023 (0.0773)	0.7879 (0.0936)	0.8822 (0.0642)
eGeMAPS (openSMILE)	RF	0	0.7291 (0.0211)	0.7184 (0.0244)	0.8000 (0.0221)	0.8215 (0.0710)	0.8119 (0.0847)	0.8979 (0.0582)
eGeMAPS (openSMILE)	LR	0	0.7483 (0.0210)	0.7413 (0.0248)	0.8142 (0.0206)	0.8397 (0.0724)	0.8283 (0.0839)	0.8839 (0.0663)

Table S2: Full binary classification performance (all four downstream classifiers per model) on Sustained Vowel tasks for the PC-GITA and NeuroVoz cohorts, at each model-classifier pair's best-performing layer (selected by segment-level $F1$ -score). This extends Table 2 in the main text, which reports only the best classifier per model. Values in parentheses are the bootstrap margin of error at a 95% confidence level.

Model	Classifier	Best Layer	Segment-level			Subject-level		
			Acc	F1	AUC	Acc	F1	AUC
PC-GITA Vowels (Colombian Spanish)								
Whisper Small	SVM	3	0.7460 (0.0227)	0.7507 (0.0246)	0.8095 (0.0225)	0.7900 (0.0784)	0.7961 (0.0841)	0.8636 (0.0702)
Whisper Small	XGB	5	0.7634 (0.0224)	0.7695 (0.0244)	0.8349 (0.0210)	0.8400 (0.0707)	0.8519 (0.0729)	0.9076 (0.0561)
Whisper Small	RF	5	0.7674 (0.0207)	0.7763 (0.0227)	0.8269 (0.0207)	0.8200 (0.0772)	0.8364 (0.0748)	0.8860 (0.0668)
Whisper Small	LR	8	0.7667 (0.0215)	0.7729 (0.0229)	0.8427 (0.0192)	0.8000 (0.0816)	0.8148 (0.0810)	0.9012 (0.0631)
Whisper Medium	SVM	4	0.7721 (0.0214)	0.7743 (0.0230)	0.8418 (0.0200)	0.8100 (0.0735)	0.8081 (0.0819)	0.9000 (0.0620)
Whisper Medium	XGB	5	0.7447 (0.0216)	0.7570 (0.0232)	0.8107 (0.0220)	0.7900 (0.0741)	0.8108 (0.0764)	0.8852 (0.0649)
Whisper Medium	RF	4	0.7540 (0.0217)	0.7638 (0.0235)	0.8348 (0.0199)	0.8000 (0.0771)	0.8113 (0.0778)	0.9036 (0.0579)
Whisper Medium	LR	4	0.7614 (0.0217)	0.7656 (0.0241)	0.8332 (0.0202)	0.8200 (0.0773)	0.8302 (0.0774)	0.8892 (0.0645)
Whisper Large	SVM	5	0.7794 (0.0201)	0.7794 (0.0224)	0.8532 (0.0189)	0.8200 (0.0745)	0.8269 (0.0780)	0.9052 (0.0622)
Whisper Large	XGB	6	0.7553 (0.0213)	0.7617 (0.0236)	0.8234 (0.0210)	0.8200 (0.0746)	0.8302 (0.0757)	0.8960 (0.0644)
Whisper Large	RF	14	0.7747 (0.0219)	0.7784 (0.0245)	0.8399 (0.0195)	0.8500 (0.0655)	0.8571 (0.0700)	0.9116 (0.0578)
Whisper Large	LR	13	0.7680 (0.0214)	0.7679 (0.0242)	0.8222 (0.0209)	0.8000 (0.0794)	0.8077 (0.0860)	0.8880 (0.0662)
Qwen3-ASR 0.6B	SVM	7	0.7674 (0.0199)	0.7620 (0.0232)	0.8336 (0.0198)	0.8400 (0.0699)	0.8367 (0.0763)	0.8792 (0.0653)
Qwen3-ASR 0.6B	XGB	8	0.7380 (0.0220)	0.7528 (0.0229)	0.8100 (0.0211)	0.8200 (0.0737)	0.8333 (0.0763)	0.8712 (0.0660)
Qwen3-ASR 0.6B	RF	7	0.7366 (0.0232)	0.7352 (0.0261)	0.8082 (0.0220)	0.8000 (0.0781)	0.8039 (0.0825)	0.8648 (0.0710)
Qwen3-ASR 0.6B	LR	2	0.7694 (0.0200)	0.7642 (0.0235)	0.8296 (0.0196)	0.8100 (0.0798)	0.8081 (0.0869)	0.8900 (0.0727)
Qwen3-ASR 1.7B	SVM	8	0.7580 (0.0205)	0.7527 (0.0237)	0.8359 (0.0194)	0.8000 (0.0763)	0.8000 (0.0851)	0.8816 (0.0643)
Qwen3-ASR 1.7B	XGB	9	0.7574 (0.0218)	0.7660 (0.0222)	0.8152 (0.0212)	0.8400 (0.0727)	0.8519 (0.0712)	0.8796 (0.0705)
Qwen3-ASR 1.7B	RF	11	0.7594 (0.0205)	0.7551 (0.0244)	0.8298 (0.0201)	0.8500 (0.0711)	0.8515 (0.0733)	0.9048 (0.0601)
Qwen3-ASR 1.7B	LR	2	0.7560 (0.0207)	0.7505 (0.0250)	0.8240 (0.0198)	0.8300 (0.0750)	0.8283 (0.0810)	0.8880 (0.0660)
Nemotron 3.5 0.6B	SVM	3	0.7761 (0.0210)	0.7678 (0.0236)	0.8478 (0.0191)	0.8300 (0.0740)	0.8283 (0.0815)	0.9080 (0.0618)
Nemotron 3.5 0.6B	XGB	10	0.7320 (0.0218)	0.7457 (0.0233)	0.8024 (0.0216)	0.8100 (0.0743)	0.8257 (0.0766)	0.8544 (0.0749)
Nemotron 3.5 0.6B	RF	10	0.7360 (0.0223)	0.7492 (0.0233)	0.8082 (0.0219)	0.7700 (0.0818)	0.7890 (0.0822)	0.8532 (0.0760)
Nemotron 3.5 0.6B	LR	11	0.7567 (0.0221)	0.7586 (0.0241)	0.8309 (0.0204)	0.7900 (0.0783)	0.7961 (0.0836)	0.8808 (0.0705)
Wav2Vec 2.0 (XLS-R)	SVM	2	0.7286 (0.0216)	0.7231 (0.0259)	0.7774 (0.0231)	0.8100 (0.0767)	0.8000 (0.0872)	0.8516 (0.0710)
Wav2Vec 2.0 (XLS-R)	XGB	9	0.6999 (0.0229)	0.6989 (0.0253)	0.7495 (0.0240)	0.7500 (0.0842)	0.7475 (0.0962)	0.8192 (0.0810)
Wav2Vec 2.0 (XLS-R)	RF	5	0.7119 (0.0220)	0.7021 (0.0270)	0.7730 (0.0231)	0.7900 (0.0759)	0.7879 (0.0852)	0.8528 (0.0714)
Wav2Vec 2.0 (XLS-R)	LR	3	0.7019 (0.0225)	0.6937 (0.0267)	0.7685 (0.0224)	0.7600 (0.0801)	0.7500 (0.0945)	0.8452 (0.0749)
OmniASR-S	SVM	2	0.7567 (0.0227)	0.7562 (0.0250)	0.8235 (0.0214)	0.8200 (0.0722)	0.8163 (0.0830)	0.9000 (0.0604)
OmniASR-S	XGB	18	0.7379 (0.0219)	0.7408 (0.0252)	0.7950 (0.0234)	0.8100 (0.0760)	0.8155 (0.0819)	0.8744 (0.0687)
OmniASR-S	RF	17	0.7365 (0.0228)	0.7453 (0.0244)	0.8195 (0.0213)	0.8300 (0.0697)	0.8411 (0.0709)	0.8968 (0.0565)
OmniASR-S	LR	17	0.7588 (0.0209)	0.7584 (0.0237)	0.8318 (0.0216)	0.8500 (0.0695)	0.8515 (0.0737)	0.8972 (0.0609)

(Supplementary Table S2 continued)

Model	Classifier	Best Layer	Segment-level			Subject-level		
			Acc	F1	AUC	Acc	F1	AUC
OmniASR-M	SVM	6	0.7433 (0.0234)	0.7424 (0.0256)	0.8176 (0.0221)	0.8100 (0.0740)	0.8119 (0.0850)	0.8812 (0.0675)
OmniASR-M	XGB	9	0.7291 (0.0232)	0.7443 (0.0239)	0.8010 (0.0225)	0.7800 (0.0825)	0.8036 (0.0804)	0.8680 (0.0708)
OmniASR-M	RF	8	0.7406 (0.0231)	0.7489 (0.0235)	0.8108 (0.0208)	0.8000 (0.0765)	0.8148 (0.0819)	0.8788 (0.0647)
OmniASR-M	LR	45	0.7507 (0.0220)	0.7563 (0.0244)	0.8179 (0.0226)	0.8000 (0.0774)	0.8113 (0.0795)	0.8772 (0.0659)
eGeMAPS (openSMILE)	SVM	0	0.7259 (0.0218)	0.7203 (0.0256)	0.7938 (0.0222)	0.8100 (0.0781)	0.8155 (0.0868)	0.8720 (0.0720)
eGeMAPS (openSMILE)	XGB	0	0.7079 (0.0219)	0.7081 (0.0262)	0.7799 (0.0214)	0.8100 (0.0751)	0.8119 (0.0827)	0.8948 (0.0625)
eGeMAPS (openSMILE)	RF	0	0.6992 (0.0223)	0.6972 (0.0264)	0.7764 (0.0219)	0.8300 (0.0718)	0.8317 (0.0783)	0.9028 (0.0577)
eGeMAPS (openSMILE)	LR	0	0.7146 (0.0218)	0.7089 (0.0255)	0.7740 (0.0224)	0.7400 (0.0909)	0.7451 (0.0983)	0.8672 (0.0713)
NeuroVoz Vowels (Castilian Spanish)								
Whisper Small	SVM	6	0.6577 (0.0294)	0.6716 (0.0333)	0.7148 (0.0299)	0.6638 (0.0893)	0.6604 (0.1050)	0.7458 (0.0980)
Whisper Small	XGB	2	0.6554 (0.0288)	0.6758 (0.0307)	0.7113 (0.0317)	0.7108 (0.0888)	0.7103 (0.0984)	0.7654 (0.0921)
Whisper Small	RF	12	0.6476 (0.0287)	0.6819 (0.0304)	0.7025 (0.0317)	0.7226 (0.0855)	0.7414 (0.0859)	0.7647 (0.0885)
Whisper Small	LR	4	0.6861 (0.0281)	0.7021 (0.0307)	0.7484 (0.0290)	0.6830 (0.0910)	0.6852 (0.1022)	0.7846 (0.0861)
Whisper Medium	SVM	10	0.6896 (0.0274)	0.7112 (0.0302)	0.7234 (0.0318)	0.6734 (0.0869)	0.6729 (0.1019)	0.7559 (0.0921)
Whisper Medium	XGB	24	0.6818 (0.0283)	0.7083 (0.0304)	0.7451 (0.0298)	0.6921 (0.0854)	0.6916 (0.0971)	0.8052 (0.0838)
Whisper Medium	RF	24	0.6858 (0.0285)	0.7113 (0.0301)	0.7389 (0.0303)	0.6932 (0.0839)	0.7027 (0.0947)	0.7965 (0.0872)
Whisper Medium	LR	18	0.7057 (0.0264)	0.7278 (0.0297)	0.7442 (0.0287)	0.7573 (0.0841)	0.7547 (0.0980)	0.7909 (0.0894)
Whisper Large	SVM	25	0.7292 (0.0284)	0.7415 (0.0289)	0.7817 (0.0298)	0.7306 (0.0873)	0.7387 (0.0990)	0.8462 (0.0748)
Whisper Large	XGB	20	0.7012 (0.0278)	0.7183 (0.0306)	0.7525 (0.0288)	0.7467 (0.0886)	0.7327 (0.1017)	0.8343 (0.0786)
Whisper Large	RF	17	0.7065 (0.0270)	0.7294 (0.0304)	0.7457 (0.0308)	0.7584 (0.0812)	0.7636 (0.0922)	0.8157 (0.0798)
Whisper Large	LR	20	0.7209 (0.0265)	0.7396 (0.0284)	0.7808 (0.0286)	0.7472 (0.0848)	0.7379 (0.0958)	0.8339 (0.0800)
Qwen3-ASR 0.6B	SVM	12	0.7182 (0.0272)	0.7321 (0.0284)	0.7777 (0.0282)	0.7766 (0.0767)	0.7778 (0.0908)	0.8434 (0.0750)
Qwen3-ASR 0.6B	XGB	12	0.7029 (0.0279)	0.7207 (0.0293)	0.7524 (0.0286)	0.7670 (0.0800)	0.7664 (0.0939)	0.8395 (0.0764)
Qwen3-ASR 0.6B	RF	12	0.7096 (0.0266)	0.7314 (0.0281)	0.7693 (0.0277)	0.7781 (0.0768)	0.7895 (0.0806)	0.8514 (0.0755)
Qwen3-ASR 0.6B	LR	12	0.7261 (0.0281)	0.7354 (0.0295)	0.7939 (0.0283)	0.7659 (0.0822)	0.7573 (0.0981)	0.8605 (0.0709)
Qwen3-ASR 1.7B	SVM	14	0.7508 (0.0254)	0.7646 (0.0277)	0.8139 (0.0252)	0.8129 (0.0739)	0.8077 (0.0843)	0.8822 (0.0633)
Qwen3-ASR 1.7B	XGB	17	0.6989 (0.0275)	0.7217 (0.0287)	0.7685 (0.0281)	0.8044 (0.0789)	0.8037 (0.0817)	0.8825 (0.0637)
Qwen3-ASR 1.7B	RF	16	0.7101 (0.0254)	0.7413 (0.0272)	0.7804 (0.0261)	0.8327 (0.0705)	0.8333 (0.0783)	0.8836 (0.0651)
Qwen3-ASR 1.7B	LR	15	0.7619 (0.0251)	0.7736 (0.0262)	0.8195 (0.0252)	0.8033 (0.0766)	0.7961 (0.0854)	0.8895 (0.0622)
Nemotron 3.5 0.6B	SVM	10	0.7688 (0.0250)	0.7833 (0.0266)	0.8365 (0.0245)	0.8418 (0.0690)	0.8411 (0.0745)	0.8990 (0.0591)
Nemotron 3.5 0.6B	XGB	12	0.7207 (0.0267)	0.7511 (0.0267)	0.7838 (0.0273)	0.7568 (0.0799)	0.7500 (0.0968)	0.8661 (0.0667)
Nemotron 3.5 0.6B	RF	16	0.7438 (0.0274)	0.7694 (0.0274)	0.7965 (0.0298)	0.8054 (0.0749)	0.8108 (0.0834)	0.8888 (0.0625)
Nemotron 3.5 0.6B	LR	9	0.7498 (0.0265)	0.7639 (0.0279)	0.8184 (0.0254)	0.7948 (0.0776)	0.7925 (0.0875)	0.8748 (0.0658)
Wav2Vec 2.0 (XLS-R)	SVM	16	0.6614 (0.0287)	0.6759 (0.0317)	0.7168 (0.0300)	0.7108 (0.0844)	0.7103 (0.0940)	0.7692 (0.0883)
Wav2Vec 2.0 (XLS-R)	XGB	10	0.6479 (0.0280)	0.6673 (0.0320)	0.7111 (0.0301)	0.7386 (0.0830)	0.7358 (0.0970)	0.8035 (0.0810)
Wav2Vec 2.0 (XLS-R)	RF	9	0.6487 (0.0293)	0.6780 (0.0311)	0.7151 (0.0302)	0.7210 (0.0844)	0.7273 (0.1016)	0.7906 (0.0849)

(Supplementary Table S2 continued)

Model	Classifier	Best Layer	Segment-level			Subject-level		
			Acc	F1	AUC	Acc	F1	AUC
Wav2Vec 2.0 (XLS-R)	LR	16	0.6876 (0.0279)	0.7011 (0.0317)	0.7446 (0.0287)	0.7472 (0.0818)	0.7379 (0.0945)	0.8063 (0.0831)
OmniASR-S	SVM	0	0.6150 (0.0318)	0.6388 (0.0339)	0.6384 (0.0355)	0.5895 (0.0948)	0.5926 (0.1106)	0.6836 (0.0987)
OmniASR-S	XGB	1	0.6275 (0.0287)	0.6465 (0.0325)	0.6552 (0.0333)	0.6750 (0.0865)	0.6903 (0.0969)	0.7357 (0.0979)
OmniASR-S	RF	14	0.5955 (0.0286)	0.6406 (0.0324)	0.6269 (0.0314)	0.6290 (0.0904)	0.6552 (0.1064)	0.6689 (0.1035)
OmniASR-S	LR	0	0.6391 (0.0282)	0.6531 (0.0335)	0.6684 (0.0334)	0.6264 (0.0922)	0.6226 (0.1080)	0.7252 (0.0976)
OmniASR-M	SVM	4	0.6007 (0.0289)	0.6131 (0.0351)	0.6376 (0.0329)	0.6082 (0.0944)	0.6111 (0.1092)	0.6808 (0.1019)
OmniASR-M	XGB	21	0.5918 (0.0285)	0.6262 (0.0318)	0.6204 (0.0337)	0.6194 (0.0866)	0.6435 (0.1014)	0.6563 (0.1019)
OmniASR-M	RF	6	0.5863 (0.0311)	0.6195 (0.0330)	0.6219 (0.0335)	0.5729 (0.0947)	0.6034 (0.1063)	0.6710 (0.1030)
OmniASR-M	LR	1	0.6047 (0.0297)	0.6271 (0.0324)	0.6213 (0.0355)	0.6178 (0.0948)	0.6239 (0.1072)	0.6801 (0.1023)
eGeMAPS (openSMILE)	SVM	0	0.6426 (0.0290)	0.6691 (0.0317)	0.6758 (0.0325)	0.6467 (0.0938)	0.6607 (0.1055)	0.7227 (0.0955)
eGeMAPS (openSMILE)	XGB	0	0.6135 (0.0284)	0.6369 (0.0320)	0.6594 (0.0317)	0.6456 (0.0951)	0.6481 (0.1058)	0.7199 (0.1006)
eGeMAPS (openSMILE)	RF	0	0.6292 (0.0283)	0.6696 (0.0309)	0.6650 (0.0318)	0.6745 (0.0921)	0.6847 (0.1046)	0.7318 (0.0949)
eGeMAPS (openSMILE)	LR	0	0.6386 (0.0296)	0.6562 (0.0325)	0.6702 (0.0329)	0.6638 (0.0908)	0.6604 (0.1075)	0.7245 (0.0936)

Table S3: Subject-level classification performance stratified by demographic subgroup (gender and age) on Sentence Reading tasks for the PC-GITA and NeuroVoz cohorts, reported for all four downstream classifiers per model at each model-classifier pair's best-performing layer (selected by segment-level $F1$ -score, identical to the layers in Supplementary Table S1). Metrics are computed on the soft-voted subject predictions of the speakers belonging to each stratum; values in parentheses are the bootstrap margin of error at a 95% confidence level. The corresponding segment-level $F1$ -scores at the same layers are plotted in Figure 7.

Model	Classifier	Best Layer	Subject-level $F1$				Subject-level ROC-AUC			
			Male	Female	Age < 65	Age ≥ 65	Male	Female	Age < 65	Age ≥ 65
PC-GITA Sentences (Colombian Spanish) — 50 male, 50 female, 64 aged < 65, 36 aged ≥ 65 subjects										
Whisper Small	SVM	6	0.8333 (0.1084)	0.8000 (0.1176)	0.8214 (0.1155)	0.8095 (0.1400)	0.9456 (0.0587)	0.8640 (0.1083)	0.9343 (0.0680)	0.8406 (0.1356)
Whisper Small	XGB	6	0.8679 (0.0984)	0.8462 (0.1029)	0.8525 (0.0972)	0.8636 (0.1155)	0.9664 (0.0401)	0.8496 (0.1225)	0.9402 (0.0606)	0.8594 (0.1383)
Whisper Small	RF	4	0.8627 (0.0967)	0.8400 (0.1031)	0.8772 (0.0898)	0.8182 (0.1282)	0.9776 (0.0305)	0.8752 (0.0975)	0.9431 (0.0649)	0.8844 (0.1110)
Whisper Small	LR	2	0.9020 (0.0842)	0.7843 (0.1201)	0.8710 (0.0900)	0.8000 (0.1421)	0.9344 (0.0792)	0.8688 (0.1106)	0.9382 (0.0665)	0.8469 (0.1378)
Whisper Medium	SVM	7	0.8846 (0.0938)	0.8235 (0.1134)	0.8852 (0.0848)	0.8095 (0.1399)	0.9632 (0.0436)	0.8896 (0.0913)	0.9422 (0.0636)	0.8750 (0.1142)
Whisper Medium	XGB	6	0.9020 (0.0898)	0.8462 (0.1058)	0.8710 (0.0900)	0.8780 (0.1087)	0.9536 (0.0576)	0.8848 (0.0990)	0.9265 (0.0685)	0.9094 (0.0956)
Whisper Medium	RF	7	0.8980 (0.0930)	0.8235 (0.1115)	0.8621 (0.0964)	0.8571 (0.1234)	0.9488 (0.0600)	0.9200 (0.0794)	0.9353 (0.0639)	0.9281 (0.0805)
Whisper Medium	LR	6	0.9057 (0.0816)	0.7917 (0.1254)	0.9000 (0.0795)	0.7805 (0.1420)	0.9616 (0.0449)	0.8752 (0.1002)	0.9490 (0.0587)	0.8375 (0.1338)
Whisper Large	SVM	14	0.8627 (0.0996)	0.8235 (0.1131)	0.8667 (0.0976)	0.8095 (0.1399)	0.9728 (0.0346)	0.8864 (0.1006)	0.9441 (0.0644)	0.8938 (0.1047)
Whisper Large	XGB	12	0.8889 (0.0982)	0.8800 (0.0951)	0.8889 (0.0893)	0.8780 (0.1081)	0.9632 (0.0456)	0.9104 (0.0838)	0.9294 (0.0720)	0.9000 (0.1089)
Whisper Large	RF	13	0.8980 (0.0889)	0.8302 (0.1060)	0.8929 (0.0839)	0.8261 (0.1215)	0.9696 (0.0363)	0.9008 (0.0881)	0.9373 (0.0632)	0.8969 (0.1006)
Whisper Large	LR	14	0.8800 (0.0941)	0.8077 (0.1141)	0.8525 (0.0956)	0.8293 (0.1354)	0.9504 (0.0511)	0.8864 (0.0981)	0.9245 (0.0722)	0.8938 (0.1039)
Qwen3-ASR 0.6B	SVM	8	0.9388 (0.0714)	0.8571 (0.1092)	0.9310 (0.0670)	0.8500 (0.1224)	0.9680 (0.0449)	0.8880 (0.1061)	0.9637 (0.0536)	0.8687 (0.1193)
Qwen3-ASR 0.6B	XGB	3	0.8889 (0.0874)	0.8000 (0.1228)	0.8475 (0.0994)	0.8444 (0.1176)	0.9440 (0.0617)	0.8640 (0.1094)	0.9343 (0.0675)	0.8344 (0.1552)
Qwen3-ASR 0.6B	RF	3	0.9057 (0.0842)	0.8085 (0.1282)	0.8772 (0.0943)	0.8372 (0.1280)	0.9904 (0.0168)	0.8560 (0.1159)	0.9363 (0.0665)	0.8469 (0.1490)
Qwen3-ASR 0.6B	LR	4	0.8846 (0.0908)	0.8077 (0.1185)	0.8852 (0.0855)	0.7907 (0.1375)	0.9584 (0.0533)	0.8640 (0.1093)	0.9431 (0.0652)	0.8500 (0.1338)
Qwen3-ASR 1.7B	SVM	9	0.9583 (0.0556)	0.8400 (0.1118)	0.9474 (0.0614)	0.8293 (0.1246)	0.9856 (0.0264)	0.8736 (0.1064)	0.9647 (0.0528)	0.8781 (0.1172)
Qwen3-ASR 1.7B	XGB	7	0.9388 (0.0714)	0.7660 (0.1290)	0.8214 (0.1134)	0.9000 (0.0986)	0.9632 (0.0484)	0.8624 (0.1002)	0.9255 (0.0743)	0.9125 (0.1001)
Qwen3-ASR 1.7B	RF	6	0.9412 (0.0645)	0.7755 (0.1245)	0.8966 (0.0839)	0.8095 (0.1360)	0.9760 (0.0353)	0.8240 (0.1251)	0.9225 (0.0812)	0.8562 (0.1298)
Qwen3-ASR 1.7B	LR	9	0.9600 (0.0532)	0.7755 (0.1269)	0.8966 (0.0839)	0.8293 (0.1246)	0.9840 (0.0268)	0.8512 (0.1097)	0.9539 (0.0569)	0.8531 (0.1279)
Nemotron 3.5 0.6B	SVM	5	0.8936 (0.0987)	0.8077 (0.1232)	0.8621 (0.0994)	0.8293 (0.1252)	0.9488 (0.0519)	0.8768 (0.1056)	0.9304 (0.0734)	0.8781 (0.1239)
Nemotron 3.5 0.6B	XGB	10	0.8571 (0.1042)	0.8302 (0.1059)	0.8475 (0.0993)	0.8372 (0.1192)	0.9504 (0.0497)	0.8624 (0.1109)	0.9039 (0.0767)	0.8781 (0.1275)
Nemotron 3.5 0.6B	RF	4	0.8980 (0.0872)	0.7925 (0.1232)	0.8136 (0.1110)	0.8837 (0.0926)	0.9392 (0.0612)	0.8800 (0.1010)	0.9010 (0.0852)	0.9250 (0.0956)
Nemotron 3.5 0.6B	LR	4	0.9200 (0.0746)	0.7843 (0.1230)	0.8667 (0.0925)	0.8293 (0.1246)	0.9600 (0.0481)	0.8400 (0.1137)	0.9167 (0.0769)	0.8594 (0.1356)
Wav2Vec 2.0 (XLS-R)	SVM	21	0.8571 (0.1046)	0.8846 (0.0900)	0.8929 (0.0839)	0.8444 (0.1176)	0.8800 (0.1036)	0.8816 (0.1041)	0.9108 (0.0819)	0.8438 (0.1330)
Wav2Vec 2.0 (XLS-R)	XGB	5	0.8627 (0.1029)	0.8333 (0.1134)	0.8772 (0.0930)	0.8095 (0.1282)	0.9136 (0.0838)	0.8800 (0.1003)	0.9216 (0.0696)	0.8594 (0.1292)
Wav2Vec 2.0 (XLS-R)	RF	5	0.8800 (0.0944)	0.8085 (0.1232)	0.8571 (0.0996)	0.8293 (0.1333)	0.9184 (0.0793)	0.8880 (0.0972)	0.9294 (0.0640)	0.8469 (0.1449)
Wav2Vec 2.0 (XLS-R)	LR	3	0.7843 (0.1276)	0.8571 (0.1078)	0.7719 (0.1270)	0.8837 (0.1069)	0.8688 (0.1041)	0.8752 (0.0985)	0.8578 (0.1054)	0.8719 (0.1168)
OmniASR-S	SVM	9	0.8511 (0.1100)	0.8235 (0.1131)	0.8571 (0.0977)	0.8095 (0.1399)	0.9168 (0.0812)	0.8800 (0.0980)	0.9098 (0.0875)	0.8625 (0.1296)
OmniASR-S	XGB	17	0.8462 (0.1077)	0.7843 (0.1194)	0.8333 (0.1040)	0.7907 (0.1375)	0.9536 (0.0476)	0.8688 (0.0968)	0.9167 (0.0726)	0.8969 (0.1025)
OmniASR-S	RF	17	0.8235 (0.1112)	0.7600 (0.1229)	0.7931 (0.1214)	0.7907 (0.1388)	0.9520 (0.0508)	0.8752 (0.0910)	0.9382 (0.0581)	0.8750 (0.1126)
OmniASR-S	LR	16	0.8571 (0.1041)	0.7843 (0.1268)	0.8000 (0.1167)	0.8500 (0.1179)	0.8960 (0.1034)	0.8480 (0.1090)	0.9059 (0.0882)	0.8250 (0.1558)
OmniASR-M	SVM	9	0.8163 (0.1162)	0.8400 (0.1066)	0.8421 (0.1046)	0.8095 (0.1261)	0.9408 (0.0566)	0.8704 (0.1031)	0.9304 (0.0686)	0.8375 (0.1373)
OmniASR-M	XGB	33	0.9091 (0.0791)	0.7200 (0.1374)	0.8387 (0.1012)	0.7907 (0.1247)	0.9504 (0.0518)	0.8512 (0.1020)	0.9265 (0.0620)	0.8812 (0.1142)
OmniASR-M	RF	32	0.8889 (0.0868)	0.8085 (0.1267)	0.8621 (0.0978)	0.8372 (0.1252)	0.9504 (0.0524)	0.8480 (0.1200)	0.9206 (0.0700)	0.8625 (0.1363)

(Supplementary Table S3 continued)

Model	Classifier	Best Layer	Subject-level F1				Subject-level ROC-AUC			
			Male	Female	Age < 65	Age ≥ 65	Male	Female	Age < 65	Age ≥ 65
OmniASR-M	LR	10	0.8750 (0.1022)	0.8235 (0.1122)	0.8621 (0.1005)	0.8293 (0.1246)	0.9360 (0.0635)	0.8816 (0.1038)	0.9206 (0.0781)	0.8688 (0.1258)
eGeMAPS (openSMILE) SVM	SVM	0	0.8679 (0.0962)	0.6667 (0.1538)	0.8000 (0.1138)	0.7317 (0.1540)	0.9376 (0.0668)	0.7312 (0.1453)	0.8618 (0.1015)	0.7500 (0.1690)
eGeMAPS (openSMILE) XGB	XGB	0	0.8421 (0.1006)	0.7083 (0.1451)	0.8333 (0.1036)	0.7111 (0.1528)	0.9408 (0.0580)	0.7616 (0.1374)	0.8735 (0.0916)	0.7531 (0.1707)
eGeMAPS (openSMILE) RF	RF	0	0.8475 (0.0983)	0.7059 (0.1427)	0.7812 (0.1166)	0.7826 (0.1311)	0.9488 (0.0575)	0.7664 (0.1322)	0.8618 (0.0978)	0.8031 (0.1623)
eGeMAPS (openSMILE) LR	LR	0	0.8627 (0.1033)	0.7347 (0.1377)	0.8197 (0.1104)	0.7692 (0.1444)	0.9040 (0.0910)	0.7792 (0.1271)	0.8588 (0.1022)	0.7969 (0.1542)
NeuroVoz Sentences (Castilian Spanish) — 61 male, 46 female, 38 aged < 65, 69 aged ≥ 65 subjects										
Whisper Small	SVM	6	0.9062 (0.0711)	0.8421 (0.1312)	0.9231 (0.1002)	0.8684 (0.0793)	0.9580 (0.0471)	0.9654 (0.0461)	0.9968 (0.0095)	0.9448 (0.0535)
Whisper Small	XGB	4	0.8710 (0.0875)	0.8500 (0.1236)	0.9231 (0.1000)	0.8421 (0.0901)	0.9192 (0.0785)	0.9308 (0.0821)	0.9903 (0.0209)	0.8846 (0.0830)
Whisper Small	RF	11	0.9231 (0.0685)	0.8649 (0.1164)	0.8462 (0.1494)	0.9211 (0.0647)	0.9504 (0.0536)	0.9615 (0.0484)	0.9643 (0.0552)	0.9525 (0.0484)
Whisper Small	LR	12	0.9375 (0.0582)	0.9500 (0.0676)	0.9231 (0.1111)	0.9487 (0.0556)	0.9720 (0.0383)	0.9904 (0.0188)	1.0000 (0.0000)	0.9745 (0.0334)
Whisper Medium	SVM	12	0.8710 (0.0875)	0.9231 (0.0897)	0.8000 (0.1831)	0.9211 (0.0610)	0.9558 (0.0474)	0.9808 (0.0313)	0.9870 (0.0301)	0.9626 (0.0404)
Whisper Medium	XGB	9	0.8525 (0.0928)	0.9000 (0.1018)	0.8333 (0.1657)	0.8831 (0.0721)	0.9407 (0.0553)	0.9731 (0.0417)	0.9740 (0.0406)	0.9414 (0.0584)
Whisper Medium	RF	18	0.9062 (0.0750)	0.8293 (0.1273)	0.8333 (0.1671)	0.8889 (0.0708)	0.9537 (0.0467)	0.9442 (0.0702)	0.9545 (0.0588)	0.9542 (0.0488)
Whisper Medium	LR	8	0.8889 (0.0840)	0.9000 (0.0961)	0.8800 (0.1429)	0.8974 (0.0682)	0.9666 (0.0442)	0.9673 (0.0445)	0.9935 (0.0159)	0.9508 (0.0503)
Whisper Large	SVM	19	0.9206 (0.0703)	0.9000 (0.0991)	0.8800 (0.1471)	0.9231 (0.0592)	0.9795 (0.0272)	0.9712 (0.0426)	0.9935 (0.0162)	0.9779 (0.0261)
Whisper Large	XGB	19	0.8889 (0.0845)	0.8500 (0.1260)	0.7826 (0.2039)	0.9000 (0.0681)	0.9634 (0.0388)	0.9288 (0.0791)	0.9383 (0.0735)	0.9686 (0.0310)
Whisper Large	RF	19	0.8923 (0.0759)	0.8500 (0.1260)	0.8000 (0.1762)	0.9000 (0.0681)	0.9537 (0.0493)	0.9250 (0.0781)	0.9318 (0.0818)	0.9559 (0.0399)
Whisper Large	LR	20	0.8889 (0.0795)	0.9048 (0.0894)	0.9231 (0.1111)	0.8861 (0.0728)	0.9784 (0.0280)	0.9635 (0.0452)	0.9935 (0.0151)	0.9669 (0.0327)
Qwen3-ASR 0.6B	SVM	10	0.9524 (0.0538)	0.9048 (0.0992)	1.0000 (0.0000)	0.9091 (0.0672)	0.9806 (0.0275)	0.9808 (0.0268)	1.0000 (0.0000)	0.9720 (0.0311)
Qwen3-ASR 0.6B	XGB	10	0.8889 (0.0821)	0.8500 (0.1244)	0.8800 (0.1471)	0.8718 (0.0783)	0.9698 (0.0365)	0.9577 (0.0558)	0.9675 (0.0484)	0.9643 (0.0353)
Qwen3-ASR 0.6B	RF	10	0.9375 (0.0592)	0.9000 (0.1048)	0.8800 (0.1471)	0.9367 (0.0564)	0.9752 (0.0344)	0.9615 (0.0540)	0.9740 (0.0429)	0.9686 (0.0374)
Qwen3-ASR 0.6B	LR	10	0.9524 (0.0538)	0.8780 (0.1094)	0.9231 (0.1088)	0.9231 (0.0614)	0.9828 (0.0243)	0.9808 (0.0286)	0.9968 (0.0095)	0.9805 (0.0244)
Qwen3-ASR 1.7B	SVM	13	0.8852 (0.0789)	0.9744 (0.0455)	0.8333 (0.1675)	0.9474 (0.0478)	0.9677 (0.0390)	1.0000 (0.0000)	1.0000 (0.0000)	0.9711 (0.0363)
Qwen3-ASR 1.7B	XGB	13	0.8852 (0.0832)	0.9189 (0.1000)	0.7826 (0.1980)	0.9333 (0.0581)	0.9720 (0.0378)	0.9923 (0.0152)	1.0000 (0.0000)	0.9779 (0.0315)
Qwen3-ASR 1.7B	RF	14	0.9206 (0.0718)	0.8649 (0.1244)	0.9231 (0.1111)	0.8919 (0.0725)	0.9731 (0.0357)	0.9769 (0.0337)	1.0000 (0.0000)	0.9626 (0.0365)
Qwen3-ASR 1.7B	LR	23	0.9062 (0.0720)	0.9231 (0.0926)	0.8800 (0.1471)	0.9231 (0.0565)	0.9741 (0.0323)	0.9827 (0.0263)	0.9935 (0.0156)	0.9796 (0.0244)
Nemotron 3.5 0.6B	SVM	10	0.9375 (0.0590)	0.9268 (0.0863)	0.9630 (0.0677)	0.9231 (0.0617)	0.9795 (0.0316)	0.9692 (0.0456)	0.9935 (0.0159)	0.9694 (0.0369)
Nemotron 3.5 0.6B	XGB	8	0.8852 (0.0879)	0.8108 (0.1462)	0.7273 (0.2164)	0.8947 (0.0726)	0.9741 (0.0360)	0.9231 (0.0788)	0.9675 (0.0516)	0.9610 (0.0387)
Nemotron 3.5 0.6B	RF	10	0.9062 (0.0731)	0.8649 (0.1228)	0.7826 (0.2039)	0.9231 (0.0616)	0.9720 (0.0381)	0.9346 (0.0798)	0.9416 (0.0828)	0.9711 (0.0368)
Nemotron 3.5 0.6B	LR	6	0.9355 (0.0595)	0.8947 (0.1023)	0.8800 (0.1500)	0.9333 (0.0607)	0.9709 (0.0344)	0.9827 (0.0265)	0.9870 (0.0273)	0.9771 (0.0282)
Wav2Vec 2.0 (XLS-R)	SVM	12	0.9375 (0.0562)	0.8947 (0.1013)	0.9630 (0.0677)	0.9067 (0.0681)	0.9741 (0.0312)	0.9808 (0.0304)	1.0000 (0.0000)	0.9643 (0.0359)
Wav2Vec 2.0 (XLS-R)	XGB	12	0.8889 (0.0826)	0.8947 (0.1065)	0.9231 (0.1111)	0.8800 (0.0769)	0.9601 (0.0418)	0.9673 (0.0458)	0.9838 (0.0295)	0.9474 (0.0482)
Wav2Vec 2.0 (XLS-R)	RF	13	0.9375 (0.0592)	0.8718 (0.1229)	0.9231 (0.1111)	0.9091 (0.0668)	0.9623 (0.0420)	0.9577 (0.0487)	0.9805 (0.0318)	0.9550 (0.0434)
Wav2Vec 2.0 (XLS-R)	LR	11	0.9394 (0.0539)	0.8718 (0.1173)	0.9231 (0.1000)	0.9114 (0.0635)	0.9892 (0.0153)	0.9673 (0.0456)	0.9935 (0.0145)	0.9694 (0.0379)
OmniASR-S	SVM	10	0.9231 (0.0650)	0.8649 (0.1238)	0.9231 (0.1002)	0.8947 (0.0716)	0.9612 (0.0402)	0.9635 (0.0461)	0.9935 (0.0143)	0.9440 (0.0468)
OmniASR-S	XGB	9	0.8387 (0.0995)	0.8205 (0.1310)	0.8462 (0.1505)	0.8267 (0.0909)	0.9343 (0.0584)	0.9635 (0.0468)	0.9903 (0.0209)	0.9168 (0.0631)
OmniASR-S	RF	17	0.9231 (0.0658)	0.8205 (0.1320)	0.9231 (0.1002)	0.8718 (0.0818)	0.9494 (0.0548)	0.9269 (0.0733)	0.9740 (0.0407)	0.9168 (0.0674)
OmniASR-S	LR	15	0.9206 (0.0744)	0.8780 (0.1081)	0.9231 (0.1002)	0.8974 (0.0680)	0.9580 (0.0509)	0.9750 (0.0337)	0.9935 (0.0143)	0.9491 (0.0546)
OmniASR-M	SVM	29	0.8889 (0.0811)	0.7895 (0.1462)	0.8333 (0.1703)	0.8571 (0.0840)	0.9731 (0.0317)	0.9327 (0.0688)	0.9870 (0.0250)	0.9397 (0.0505)
OmniASR-M	XGB	19	0.8852 (0.0853)	0.8947 (0.1080)	0.8800 (0.1471)	0.8919 (0.0729)	0.9806 (0.0258)	0.9731 (0.0379)	1.0000 (0.0000)	0.9754 (0.0266)

(Supplementary Table S3 continued)

Model	Classifier	Best Layer	Subject-level $F1$				Subject-level ROC-AUC			
			Male	Female	Age < 65	Age $\geq$ 65	Male	Female	Age < 65	Age $\geq$ 65
OmniASR-M	RF	30	0.8824 (0.0769)	0.8947 (0.1013)	0.8571 (0.1551)	0.8974 (0.0780)	0.9440 (0.0571)	0.9558 (0.0578)	0.9513 (0.0625)	0.9516 (0.0508)
OmniASR-M	LR	10	0.9231 (0.0690)	0.9268 (0.0854)	0.9630 (0.0714)	0.9114 (0.0655)	0.9580 (0.0517)	0.9673 (0.0420)	0.9903 (0.0208)	0.9457 (0.0537)
eGeMAPS (openSMILE) SVM		0	0.7667 (0.1184)	0.8095 (0.1282)	0.8148 (0.1602)	0.7733 (0.1039)	0.8384 (0.1024)	0.9038 (0.0856)	0.9221 (0.0852)	0.8217 (0.0990)
eGeMAPS (openSMILE) XGB		0	0.7667 (0.1184)	0.8205 (0.1333)	0.8462 (0.1494)	0.7671 (0.1069)	0.8481 (0.0972)	0.9288 (0.0755)	0.9578 (0.0557)	0.8226 (0.0990)
eGeMAPS (openSMILE) RF		0	0.7869 (0.1190)	0.8500 (0.1214)	0.8462 (0.1494)	0.8000 (0.1003)	0.8675 (0.0904)	0.9404 (0.0740)	0.9578 (0.0568)	0.8531 (0.0890)
eGeMAPS (openSMILE) LR		0	0.8197 (0.1047)	0.8421 (0.1289)	0.8462 (0.1494)	0.8219 (0.0940)	0.8696 (0.0893)	0.9077 (0.0893)	0.9740 (0.0391)	0.8260 (0.0992)

Table S4: Subject-level classification performance stratified by demographic subgroup (gender and age) on Sustained Vowel tasks for the PC-GITA and NeuroVoz cohorts, reported for all four downstream classifiers per model at each model-classifier pair's best-performing layer (selected by segment-level $F1$ -score, identical to the layers in Supplementary Table S2). Metrics are computed on the soft-voted subject predictions of the speakers belonging to each stratum; values in parentheses are the bootstrap margin of error at a 95% confidence level. The corresponding segment-level $F1$ -scores at the same layers are plotted in Supplementary Figure S6.

Model	Classifier	Best Layer	Subject-level F1				Subject-level ROC-AUC			
			Male	Female	Age < 65	Age ≥ 65	Male	Female	Age < 65	Age ≥ 65
PC-GITA Vowels (Colombian Spanish) — 50 male, 50 female, 64 aged < 65, 36 aged ≥ 65 subjects										
Whisper Small	SVM	3	0.8519 (0.1013)	0.7347 (0.1461)	0.7937 (0.1111)	0.8000 (0.1322)	0.9024 (0.0833)	0.7936 (0.1188)	0.8471 (0.0978)	0.8844 (0.1083)
Whisper Small	XGB	5	0.8679 (0.0969)	0.8364 (0.1043)	0.8485 (0.0917)	0.8571 (0.1171)	0.9296 (0.0721)	0.8800 (0.0916)	0.9147 (0.0675)	0.9031 (0.1030)
Whisper Small	RF	5	0.8519 (0.0990)	0.8214 (0.1082)	0.8485 (0.0925)	0.8182 (0.1226)	0.9056 (0.0839)	0.8544 (0.1055)	0.8794 (0.0854)	0.8875 (0.1106)
Whisper Small	LR	8	0.8679 (0.0946)	0.7636 (0.1227)	0.7812 (0.1067)	0.8636 (0.1124)	0.9120 (0.0823)	0.8784 (0.0919)	0.8735 (0.0840)	0.9563 (0.0636)
Whisper Medium	SVM	4	0.8627 (0.0964)	0.7500 (0.1397)	0.8000 (0.1112)	0.8205 (0.1248)	0.9200 (0.0846)	0.8592 (0.1035)	0.8931 (0.0764)	0.9219 (0.0882)
Whisper Medium	XGB	5	0.8727 (0.0927)	0.7500 (0.1252)	0.8000 (0.1038)	0.8261 (0.1201)	0.9088 (0.0880)	0.8432 (0.1079)	0.9059 (0.0708)	0.8594 (0.1272)
Whisper Medium	RF	4	0.8889 (0.0881)	0.7308 (0.1369)	0.7812 (0.1112)	0.8571 (0.1137)	0.9408 (0.0660)	0.8448 (0.1035)	0.9010 (0.0750)	0.9094 (0.0926)
Whisper Medium	LR	4	0.8846 (0.0898)	0.7778 (0.1202)	0.8125 (0.1016)	0.8571 (0.1089)	0.9088 (0.0945)	0.8432 (0.1085)	0.8892 (0.0789)	0.8844 (0.1239)
Whisper Large	SVM	5	0.8679 (0.0996)	0.7843 (0.1260)	0.8125 (0.1037)	0.8500 (0.1148)	0.9104 (0.0912)	0.8800 (0.0915)	0.9039 (0.0745)	0.9094 (0.0984)
Whisper Large	XGB	6	0.8462 (0.1034)	0.8148 (0.1107)	0.8308 (0.1004)	0.8293 (0.1287)	0.8688 (0.1133)	0.9040 (0.0825)	0.9020 (0.0789)	0.8906 (0.1088)
Whisper Large	RF	14	0.8846 (0.0913)	0.8302 (0.1044)	0.8254 (0.1035)	0.9048 (0.0894)	0.9232 (0.0875)	0.8816 (0.0932)	0.8990 (0.0769)	0.9250 (0.0969)
Whisper Large	LR	13	0.8400 (0.1100)	0.7778 (0.1168)	0.8065 (0.1073)	0.8095 (0.1355)	0.9120 (0.0990)	0.8544 (0.0986)	0.8873 (0.0873)	0.8812 (0.1245)
Qwen3-ASR 0.6B	SVM	7	0.8800 (0.0951)	0.7917 (0.1254)	0.8571 (0.0978)	0.8095 (0.1334)	0.9184 (0.0912)	0.8224 (0.1176)	0.8971 (0.0829)	0.8375 (0.1539)
Qwen3-ASR 0.6B	XGB	8	0.9020 (0.0830)	0.7719 (0.1199)	0.8125 (0.1027)	0.8636 (0.1096)	0.9152 (0.0857)	0.8128 (0.1146)	0.8814 (0.0856)	0.8250 (0.1516)
Qwen3-ASR 0.6B	RF	7	0.8800 (0.0951)	0.7308 (0.1349)	0.7797 (0.1131)	0.8372 (0.1181)	0.9040 (0.0975)	0.8096 (0.1153)	0.8902 (0.0820)	0.8094 (0.1573)
Qwen3-ASR 0.6B	LR	2	0.8000 (0.1183)	0.8163 (0.1201)	0.7931 (0.1089)	0.8293 (0.1252)	0.9344 (0.0687)	0.8448 (0.1167)	0.8902 (0.0871)	0.8969 (0.1064)
Qwen3-ASR 1.7B	SVM	8	0.8846 (0.0894)	0.7083 (0.1372)	0.7931 (0.1207)	0.8095 (0.1334)	0.9392 (0.0660)	0.8128 (0.1182)	0.8843 (0.0880)	0.8469 (0.1303)
Qwen3-ASR 1.7B	XGB	9	0.8846 (0.0906)	0.8214 (0.1092)	0.8571 (0.0964)	0.8444 (0.1116)	0.9344 (0.0674)	0.8224 (0.1212)	0.8980 (0.0831)	0.8406 (0.1313)
Qwen3-ASR 1.7B	RF	11	0.8980 (0.0918)	0.8077 (0.1228)	0.8364 (0.1095)	0.8696 (0.1042)	0.9552 (0.0617)	0.8432 (0.1151)	0.9304 (0.0619)	0.8438 (0.1500)
Qwen3-ASR 1.7B	LR	2	0.8800 (0.0944)	0.7755 (0.1254)	0.8475 (0.0993)	0.8000 (0.1413)	0.9088 (0.0871)	0.8496 (0.1093)	0.9078 (0.0772)	0.8500 (0.1303)
Nemotron 3.5 0.6B	SVM	3	0.8800 (0.0954)	0.7755 (0.1370)	0.8214 (0.1115)	0.8372 (0.1181)	0.9520 (0.0601)	0.8576 (0.1095)	0.9216 (0.0748)	0.8750 (0.1257)
Nemotron 3.5 0.6B	XGB	10	0.8846 (0.0902)	0.7719 (0.1186)	0.8000 (0.1029)	0.8636 (0.1125)	0.9280 (0.0702)	0.7680 (0.1306)	0.8549 (0.0938)	0.8562 (0.1276)
Nemotron 3.5 0.6B	RF	10	0.8400 (0.1067)	0.7458 (0.1209)	0.7879 (0.1046)	0.7907 (0.1364)	0.9136 (0.0852)	0.7872 (0.1245)	0.8618 (0.0947)	0.8469 (0.1384)
Nemotron 3.5 0.6B	LR	11	0.8980 (0.0889)	0.7037 (0.1373)	0.7742 (0.1113)	0.8293 (0.1246)	0.9504 (0.0633)	0.7936 (0.1201)	0.8667 (0.0908)	0.9125 (0.1048)
Wav2Vec 2.0 (XLS-R)	SVM	2	0.7660 (0.1329)	0.8333 (0.1181)	0.7719 (0.1230)	0.8421 (0.1259)	0.7920 (0.1229)	0.9216 (0.0757)	0.8451 (0.1072)	0.8719 (0.1210)
Wav2Vec 2.0 (XLS-R)	XGB	9	0.6667 (0.1501)	0.8235 (0.1189)	0.7241 (0.1358)	0.7805 (0.1387)	0.7664 (0.1215)	0.8576 (0.1040)	0.8059 (0.1088)	0.8313 (0.1310)
Wav2Vec 2.0 (XLS-R)	RF	5	0.7660 (0.1328)	0.8077 (0.1212)	0.7368 (0.1308)	0.8571 (0.1173)	0.8480 (0.1019)	0.8672 (0.1002)	0.8304 (0.0983)	0.8969 (0.1206)
Wav2Vec 2.0 (XLS-R)	LR	3	0.6939 (0.1472)	0.8085 (0.1257)	0.7636 (0.1316)	0.7317 (0.1527)	0.7904 (0.1247)	0.9120 (0.0732)	0.8431 (0.1003)	0.8406 (0.1346)
OmniASR-S	SVM	2	0.8627 (0.1078)	0.7660 (0.1323)	0.7586 (0.1330)	0.9000 (0.0942)	0.9040 (0.0944)	0.8736 (0.0938)	0.8922 (0.0761)	0.9031 (0.1095)
OmniASR-S	XGB	18	0.8846 (0.0933)	0.7451 (0.1335)	0.8254 (0.1024)	0.8000 (0.1413)	0.8896 (0.1058)	0.8224 (0.1164)	0.8775 (0.0881)	0.8625 (0.1379)
OmniASR-S	RF	17	0.9020 (0.0899)	0.7857 (0.1179)	0.8254 (0.1035)	0.8636 (0.1099)	0.9360 (0.0756)	0.8432 (0.1063)	0.8941 (0.0757)	0.8875 (0.1144)
OmniASR-S	LR	17	0.8846 (0.0937)	0.8163 (0.1186)	0.8333 (0.1055)	0.8780 (0.1064)	0.9344 (0.0704)	0.8448 (0.1142)	0.8941 (0.0834)	0.8844 (0.1220)
OmniASR-M	SVM	6	0.8462 (0.1123)	0.7755 (0.1331)	0.7458 (0.1290)	0.9048 (0.0954)	0.8848 (0.1093)	0.8704 (0.0971)	0.8608 (0.0945)	0.8969 (0.1145)
OmniASR-M	XGB	9	0.8679 (0.0978)	0.7458 (0.1222)	0.7761 (0.1133)	0.8444 (0.1218)	0.9136 (0.0936)	0.7952 (0.1166)	0.8373 (0.0991)	0.9219 (0.0922)
OmniASR-M	RF	8	0.9057 (0.0830)	0.7273 (0.1276)	0.7879 (0.1111)	0.8571 (0.1162)	0.9120 (0.0846)	0.8224 (0.1080)	0.8618 (0.0860)	0.9156 (0.0950)

(Supplementary Table S4 continued)

Model	Classifier	Best Layer	Subject-level F1				Subject-level ROC-AUC			
			Male	Female	Age < 65	Age ≥ 65	Male	Female	Age < 65	Age ≥ 65
OmniASR-M	LR	45	0.8235 (0.1102)	0.8000 (0.1167)	0.7937 (0.1128)	0.8372 (0.1265)	0.9104 (0.0913)	0.8512 (0.1085)	0.8745 (0.0871)	0.8781 (0.1266)
eGeMAPS (openSMILE) SVM	SVM	0	0.8235 (0.1086)	0.8077 (0.1156)	0.7742 (0.1136)	0.8780 (0.1122)	0.9072 (0.0855)	0.8336 (0.1141)	0.8333 (0.0969)	0.9219 (0.0988)
eGeMAPS (openSMILE) XGB	XGB	0	0.8400 (0.1111)	0.7843 (0.1264)	0.7869 (0.1128)	0.8500 (0.1224)	0.9168 (0.0877)	0.8544 (0.0979)	0.8833 (0.0778)	0.9031 (0.1130)
eGeMAPS (openSMILE) RF	RF	0	0.8800 (0.0975)	0.7843 (0.1221)	0.8197 (0.1055)	0.8500 (0.1202)	0.9280 (0.0816)	0.8768 (0.0886)	0.8843 (0.0790)	0.9219 (0.0988)
eGeMAPS (openSMILE) LR	LR	0	0.7600 (0.1330)	0.7308 (0.1380)	0.6885 (0.1285)	0.8293 (0.1243)	0.8992 (0.0821)	0.8480 (0.1019)	0.8216 (0.1059)	0.9469 (0.0641)
NeuroVoz Vowels (Castilian Spanish) — 61 male, 46 female, 38 aged < 65, 69 aged ≥ 65 subjects										
Whisper Small	SVM	6	0.6462 (0.1328)	0.6829 (0.1630)	0.5385 (0.2363)	0.7000 (0.1209)	0.7522 (0.1215)	0.7596 (0.1504)	0.8117 (0.1628)	0.6766 (0.1323)
Whisper Small	XGB	2	0.7097 (0.1209)	0.7111 (0.1518)	0.7586 (0.1783)	0.6923 (0.1205)	0.7834 (0.1201)	0.7558 (0.1461)	0.8474 (0.1153)	0.6834 (0.1295)
Whisper Small	RF	12	0.7761 (0.1111)	0.6939 (0.1485)	0.7586 (0.1818)	0.7356 (0.1059)	0.7737 (0.1240)	0.7635 (0.1492)	0.8636 (0.1218)	0.6893 (0.1371)
Whisper Small	LR	4	0.6452 (0.1401)	0.7391 (0.1491)	0.6667 (0.2126)	0.6914 (0.1180)	0.8028 (0.1112)	0.7769 (0.1458)	0.8539 (0.1138)	0.7275 (0.1215)
Whisper Medium	SVM	10	0.7000 (0.1446)	0.6383 (0.1586)	0.6154 (0.2245)	0.6914 (0.1190)	0.7920 (0.1159)	0.7327 (0.1539)	0.8052 (0.1585)	0.6961 (0.1313)
Whisper Medium	XGB	24	0.7119 (0.1328)	0.6667 (0.1600)	0.5926 (0.2164)	0.7250 (0.1124)	0.8373 (0.1029)	0.7904 (0.1476)	0.7825 (0.1539)	0.7878 (0.1121)
Whisper Medium	RF	24	0.6984 (0.1346)	0.7083 (0.1517)	0.6154 (0.2244)	0.7294 (0.1102)	0.8136 (0.1103)	0.7942 (0.1509)	0.7890 (0.1575)	0.7750 (0.1158)
Whisper Medium	LR	18	0.7302 (0.1318)	0.7907 (0.1408)	0.7692 (0.1888)	0.7595 (0.1081)	0.7942 (0.1172)	0.8000 (0.1458)	0.8701 (0.1325)	0.7292 (0.1230)
Whisper Large	SVM	25	0.7188 (0.1145)	0.7660 (0.1376)	0.6897 (0.2026)	0.7561 (0.1048)	0.8427 (0.0952)	0.8615 (0.1232)	0.8442 (0.1453)	0.8124 (0.1016)
Whisper Large	XGB	20	0.7119 (0.1271)	0.7619 (0.1469)	0.5833 (0.2500)	0.7792 (0.1020)	0.8233 (0.1035)	0.8673 (0.1063)	0.8084 (0.1420)	0.8379 (0.0951)
Whisper Large	RF	17	0.7692 (0.1147)	0.7556 (0.1402)	0.7857 (0.1758)	0.7561 (0.1081)	0.8039 (0.1244)	0.8462 (0.1199)	0.8864 (0.1059)	0.7683 (0.1145)
Whisper Large	LR	20	0.7213 (0.1204)	0.7619 (0.1412)	0.6923 (0.2257)	0.7532 (0.1121)	0.8265 (0.1029)	0.8538 (0.1142)	0.8669 (0.1238)	0.8056 (0.1041)
Qwen3-ASR 0.6B	SVM	12	0.8065 (0.1059)	0.7391 (0.1416)	0.6154 (0.2289)	0.8293 (0.0919)	0.8588 (0.0925)	0.8192 (0.1306)	0.8279 (0.1315)	0.8421 (0.1009)
Qwen3-ASR 0.6B	XGB	12	0.7937 (0.1111)	0.7273 (0.1451)	0.6667 (0.2126)	0.8000 (0.0933)	0.8448 (0.1004)	0.8442 (0.1128)	0.8474 (0.1222)	0.8294 (0.1028)
Qwen3-ASR 0.6B	RF	12	0.8182 (0.1045)	0.7500 (0.1393)	0.6897 (0.1905)	0.8235 (0.0900)	0.8502 (0.0992)	0.8654 (0.1117)	0.8214 (0.1378)	0.8523 (0.0994)
Qwen3-ASR 0.6B	LR	12	0.7869 (0.1128)	0.7143 (0.1548)	0.6400 (0.2263)	0.7949 (0.1027)	0.8664 (0.0925)	0.8596 (0.1150)	0.8344 (0.1367)	0.8633 (0.0958)
Qwen3-ASR 1.7B	SVM	14	0.8387 (0.1002)	0.7619 (0.1464)	0.7407 (0.1926)	0.8312 (0.0891)	0.8976 (0.0780)	0.8731 (0.1022)	0.9091 (0.0938)	0.8591 (0.0905)
Qwen3-ASR 1.7B	XGB	17	0.8525 (0.0992)	0.7391 (0.1527)	0.7407 (0.1926)	0.8250 (0.0956)	0.9267 (0.0697)	0.8365 (0.1169)	0.8864 (0.1089)	0.8727 (0.0935)
Qwen3-ASR 1.7B	RF	16	0.8438 (0.0946)	0.8182 (0.1348)	0.7407 (0.1825)	0.8642 (0.0824)	0.9041 (0.0797)	0.8750 (0.1152)	0.8961 (0.1078)	0.8650 (0.0935)
Qwen3-ASR 1.7B	LR	15	0.8197 (0.1037)	0.7619 (0.1445)	0.7407 (0.1907)	0.8158 (0.0919)	0.8966 (0.0803)	0.8904 (0.0899)	0.9026 (0.1002)	0.8727 (0.0866)
Nemotron 3.5 0.6B	SVM	10	0.8182 (0.1016)	0.8780 (0.1095)	0.7857 (0.1726)	0.8608 (0.0830)	0.8815 (0.0892)	0.9058 (0.0978)	0.8799 (0.1270)	0.8930 (0.0796)
Nemotron 3.5 0.6B	XGB	12	0.7869 (0.1111)	0.6977 (0.1566)	0.7407 (0.1851)	0.7532 (0.1080)	0.8772 (0.0880)	0.8615 (0.1062)	0.8831 (0.1073)	0.8463 (0.0950)
Nemotron 3.5 0.6B	RF	16	0.8438 (0.0961)	0.7660 (0.1350)	0.7407 (0.1983)	0.8333 (0.0847)	0.9106 (0.0725)	0.8731 (0.1066)	0.8701 (0.1205)	0.8913 (0.0798)
Nemotron 3.5 0.6B	LR	9	0.8254 (0.1043)	0.7442 (0.1475)	0.8148 (0.1739)	0.7848 (0.1034)	0.8782 (0.0939)	0.8769 (0.1052)	0.8961 (0.1063)	0.8497 (0.0918)
Wav2Vec 2.0 (XLS-R)	SVM	16	0.7188 (0.1245)	0.6977 (0.1591)	0.6087 (0.2407)	0.7381 (0.1024)	0.7888 (0.1170)	0.7404 (0.1519)	0.7760 (0.1585)	0.7334 (0.1199)
Wav2Vec 2.0 (XLS-R)	XGB	10	0.7213 (0.1249)	0.7556 (0.1386)	0.6400 (0.2194)	0.7654 (0.1043)	0.8093 (0.1044)	0.8192 (0.1220)	0.8312 (0.1302)	0.7733 (0.1120)
Wav2Vec 2.0 (XLS-R)	RF	9	0.7213 (0.1214)	0.7347 (0.1424)	0.6667 (0.2057)	0.7470 (0.1030)	0.8071 (0.1089)	0.7865 (0.1279)	0.8052 (0.1410)	0.7572 (0.1146)
Wav2Vec 2.0 (XLS-R)	LR	16	0.7541 (0.1215)	0.7143 (0.1589)	0.6400 (0.2262)	0.7692 (0.1088)	0.8254 (0.1068)	0.7808 (0.1390)	0.7727 (0.1687)	0.7861 (0.1068)
OmniASR-S	SVM	0	0.6207 (0.1548)	0.5600 (0.1614)	0.5000 (0.2250)	0.6250 (0.1255)	0.7446 (0.1254)	0.6000 (0.1723)	0.7143 (0.1863)	0.6188 (0.1398)
OmniASR-S	XGB	1	0.6970 (0.1220)	0.6809 (0.1516)	0.5714 (0.2194)	0.7294 (0.1100)	0.7425 (0.1230)	0.7404 (0.1414)	0.7240 (0.1703)	0.7250 (0.1277)
OmniASR-S	RF	14	0.6875 (0.1338)	0.6154 (0.1497)	0.6667 (0.2159)	0.6517 (0.1150)	0.7478 (0.1284)	0.5904 (0.1705)	0.7630 (0.1550)	0.5866 (0.1383)
OmniASR-S	LR	0	0.6786 (0.1509)	0.5600 (0.1655)	0.5385 (0.2204)	0.6500 (0.1194)	0.7996 (0.1134)	0.6250 (0.1715)	0.7532 (0.1563)	0.6715 (0.1341)
OmniASR-M	SVM	4	0.6207 (0.1449)	0.6000 (0.1607)	0.4800 (0.2352)	0.6506 (0.1173)	0.7091 (0.1311)	0.6462 (0.1591)	0.6883 (0.1766)	0.6350 (0.1308)
OmniASR-M	XGB	21	0.6769 (0.1299)	0.6000 (0.1534)	0.6429 (0.2092)	0.6437 (0.1196)	0.7069 (0.1387)	0.5827 (0.1652)	0.7468 (0.1617)	0.5713 (0.1485)

(Supplementary Table S4 continued)

Model	Classifier	Best Layer	Subject-level F1				Subject-level ROC-AUC			
			Male	Female	Age < 65	Age ≥ 65	Male	Female	Age < 65	Age ≥ 65
OmniASR-M	RF	6	0.6769 (0.1292)	0.5098 (0.1631)	0.4828 (0.2285)	0.6437 (0.1215)	0.7543 (0.1251)	0.5327 (0.1696)	0.6299 (0.1840)	0.6613 (0.1334)
OmniASR-M	LR	1	0.6333 (0.1443)	0.6122 (0.1571)	0.5161 (0.2143)	0.6667 (0.1256)	0.7371 (0.1295)	0.6038 (0.1691)	0.6299 (0.1795)	0.6715 (0.1304)
eGeMAPS (openSMILE) SVM		0	0.6452 (0.1412)	0.6800 (0.1476)	0.6154 (0.2310)	0.6744 (0.1109)	0.7403 (0.1257)	0.7462 (0.1500)	0.7825 (0.1424)	0.6596 (0.1343)
eGeMAPS (openSMILE) XGB		0	0.6452 (0.1324)	0.6522 (0.1649)	0.5185 (0.2386)	0.6914 (0.1206)	0.7274 (0.1275)	0.7288 (0.1569)	0.7435 (0.1564)	0.6868 (0.1276)
eGeMAPS (openSMILE) RF		0	0.7097 (0.1297)	0.6531 (0.1626)	0.5926 (0.2282)	0.7143 (0.1143)	0.7726 (0.1224)	0.6981 (0.1582)	0.7890 (0.1475)	0.6859 (0.1304)
eGeMAPS (openSMILE) LR		0	0.6780 (0.1440)	0.6383 (0.1636)	0.5833 (0.2390)	0.6829 (0.1181)	0.7651 (0.1266)	0.6981 (0.1568)	0.7338 (0.1611)	0.6851 (0.1293)

Table S5: NeuroVoz sentences, English translations, and assessment-task category, as documented for the NeuroVoz corpus [1, 2]. These are the 16 phonetically balanced sentences referenced as the NeuroVoz sentence-reading task in the main text.

#	Spanish Sentence	English Translation	Task Category
1	La patata no está bien ablandada	The potato is not soft enough	Velopharyngeal Closure
2	Mañana vamos de acampada	Tomorrow we are going camping	Velopharyngeal Closure
3	Cuando las barbas de tu vecino veas pelar, pon las tuyas a remojar	When your neighbor's beard you see peeling, put yours to soak	Prosody
4	Burro grande ande o no ande	Big donkey walk or not walk	Intonation-Emotion
5	De la calle vendrá quien de tu casa te echará	From outside will come that who will kick you out from your house	Intonation-Emotion
6	Carmen baila el mambo	Carmen dances the mambo	Velopharyngeal Closure
7	Cuando el diablo no sabe qué hacer, con el rabo mata moscas	When the devil does not know what to do, it kills flies with its tail	Prosody
8	Esto es una ganga	This is a bargain	Intonation-Emotion
9	Juan tira de la manga	Juan pulls the sleeve	Velopharyngeal Closure
10	Dame pan y llámame perro	Give me bread and call me dog	Velopharyngeal Closure
11	Al pan pan y al vino vino	To the bread bread and to the wine wine	Velopharyngeal Closure
12	La patata blanda es buena	The soft potato is good	Velopharyngeal Closure
13	La petaca blanca es mía	The white flask is mine	Velopharyngeal Closure
14	No pidas a quien pidió, ni sirvas a quien sirvió	Do not beg the one who begged, nor serve the person who served	Intonation-Emotion
15	El que a buen árbol se arrima, buena sombra le cobija	Who leans close to a good tree is sheltered by good shade	Intonation-Emotion
16	Tomás tira de la manta	Tomás pulls the blanket	Intonation-Emotion

Table S6: PC-GITA sentence-repetition prompts, comprising sentences varying in syntactic complexity (simple vs. complex) and sentences with word-level emphasis (capitalized) for prosodic-stress assessment, as documented for the PC-GITA corpus [3]. Capitalized words in the Emphasized set indicate the syllables/words participants were asked to stress.

#	Type	Spanish Sentence	English Translation
Sentence Repetition: Simple vs. Complex Syntax			
a	Simple	Mi casa tiene tres cuartos.	My house has three rooms.
b	Complex	Omar, que vive cerca, trajo miel.	Omar, who lives nearby, brought honey.
c	Complex	Laura sube al tren que pasa.	Laura gets on the train that is passing by.
d	Simple	Los libros nuevos no caben en la mesa de la oficina.	The new books don't fit on the office table.
e	Complex	Rosita Niño, que pinta bien, donó sus cuadros ayer.	Rosita Niño, who is a good painter, donated her artworks yesterday.
f	Complex	Luisa Rey compra el colchón duro que tanto le gusta.	Luisa Rey buys the firm mattress that she likes so much.
Emphasized Sentence Reading			
a	Emphasized	¿Viste las noticias? Yo vi GANAR la medalla de plata en pesas. ¡Ese muchacho tiene mucha fuerza!	Did you see the news? I saw him WIN the silver medal in weightlifting. That guy is really strong!
b	Emphasized	Juan se ROMPIÓ una PIERNA cuando iba en la MOTO.	Juan BROKE his LEG while riding the MOTORCYCLE.
c	Emphasized	Estoy muy triste, ayer vi MORIR a un amigo.	I'm very sad — yesterday I saw a friend DIE.
d	Emphasized	Estoy muy preocupado, cada vez me es más difícil HABLAR.	I'm very worried — it's getting harder and harder to TALK.

Table S7: Overview of the 88 eGeMAPS acoustic features and their functionals, extracted with openSMILE [4, 5]. This extends the brief feature summary given in the Methods.

Domain (# Features)	Feature	Description
Frequency (8)	Pitch (F_0)	Logarithmic fundamental frequency
	Jitter	Deviations in individual consecutive F_0 period lengths
	Formant frequency (3)	1–3 Center frequency of the 1st–3rd formant
	Formant bandwidth (3)	1–3 Bandwidth of the 1st–3rd formant
Energy (4)	Shimmer	Difference of the peak amplitudes of consecutive F_0 periods
	Loudness	Estimated perceived signal intensity from an auditory spectrum
	HNR	Ratio of energy in harmonic components to energy in noise-like components
	Equivalent sound level	Sound level for the noise environment
Spectral (14)	Alpha ratio	Spectral balance between low- and high-frequency components
	Hammarberg index	Height of the highest peak in the cepstrum relative to the surrounding peaks
	Spectral slope (0–500/500–1500 Hz) (2)	Rate of change of spectral energy distribution across frequencies
	Formant 1–3 relative energy (3)	Ratio of energy of the spectral harmonic peak at each formant’s center
	Harmonic difference H1–H2, H1–A3 (2)	Ratio of energy between harmonics
	MFCC 1–4 (4)	Mel-frequency cepstral coefficients
	Spectral flux	Difference between the spectra of two consecutive frames
Temporal (4)	Rate of loudness peaks	Number of loudness peaks per second
	Rate of voiced segments	Number of continuous voiced regions per second ($F_0 > 0$)
	Length of voiced segments	Duration of continuous voiced regions
	Length of unvoiced segments	Duration of continuous unvoiced regions ($F_0 = 0$)
Functionals applied to eligible low-level descriptors		
	Arithmetic mean / standard deviation	Average and variability of a set of feature values
	Coefficient of variation	Relative variability, expressed as a percentage of the mean
	20th/80th percentile	Values at the 20th and 80th percentiles
	Range of 20th–80th percentile	Range spanned between the 20th and 80th percentiles
	Rising/falling slope	Slope of the rising/falling segment of the contour

Table S8: Hyperparameter configurations for the downstream classifiers evaluated in this study. No explicit hyperparameter grid search or tuning was conducted; parameters were selected empirically for baseline stability across models and tasks.

Classifier	Hyperparameter	Value / Setting in Pipeline
Support Vector Machine (SVM)	Kernel	Radial Basis Function (RBF)
	Regularization parameter (C)	1.0
	Kernel coefficient (γ)	Scale ($1/(d \times \text{Var}(X))$)
	Class weights	Balanced
	Probability estimation	Enabled (Platt scaling)
XGBoost (XGB)	Objective	Binary logistic ('binary:logistic')
	Evaluation metric	Logarithmic loss ('logloss')
	Learning rate (η)	0.05
	Maximum tree depth	6
	Number of estimators	100
Random Forest (RF)	Number of estimators	100
	Maximum tree depth	Unlimited ('None')
	Minimum samples to split	5
	Minimum samples at leaf	2
	Maximum features per split	Square root of total features ($\sqrt{d}$)
	Class weights	Balanced
Logistic Regression (LR)	Regularization penalty	L_2 regularization
	Regularization strength (C)	1.0
	Optimization solver	L-BFGS ('lbfgs')
	Maximum iterations	1000
	Class weights	Balanced

Supplementary Figures

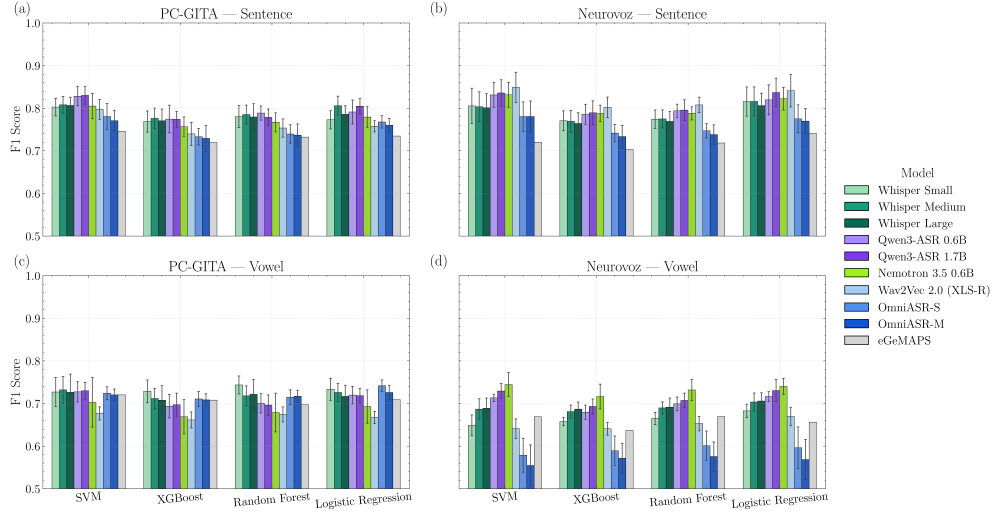


Fig. S1 Comparison of segment-level F1-scores across pre-trained foundation-model representations and handcrafted acoustic parameters for the PC-GITA sentence reading tasks (a), NeuroVoz sentence reading tasks (b), PC-GITA sustained vowel tasks (c), and NeuroVoz sustained vowel tasks (d). Bars summarize scores across evaluated layers for each model-classifier combination, and error bars denote the standard deviation across layers.

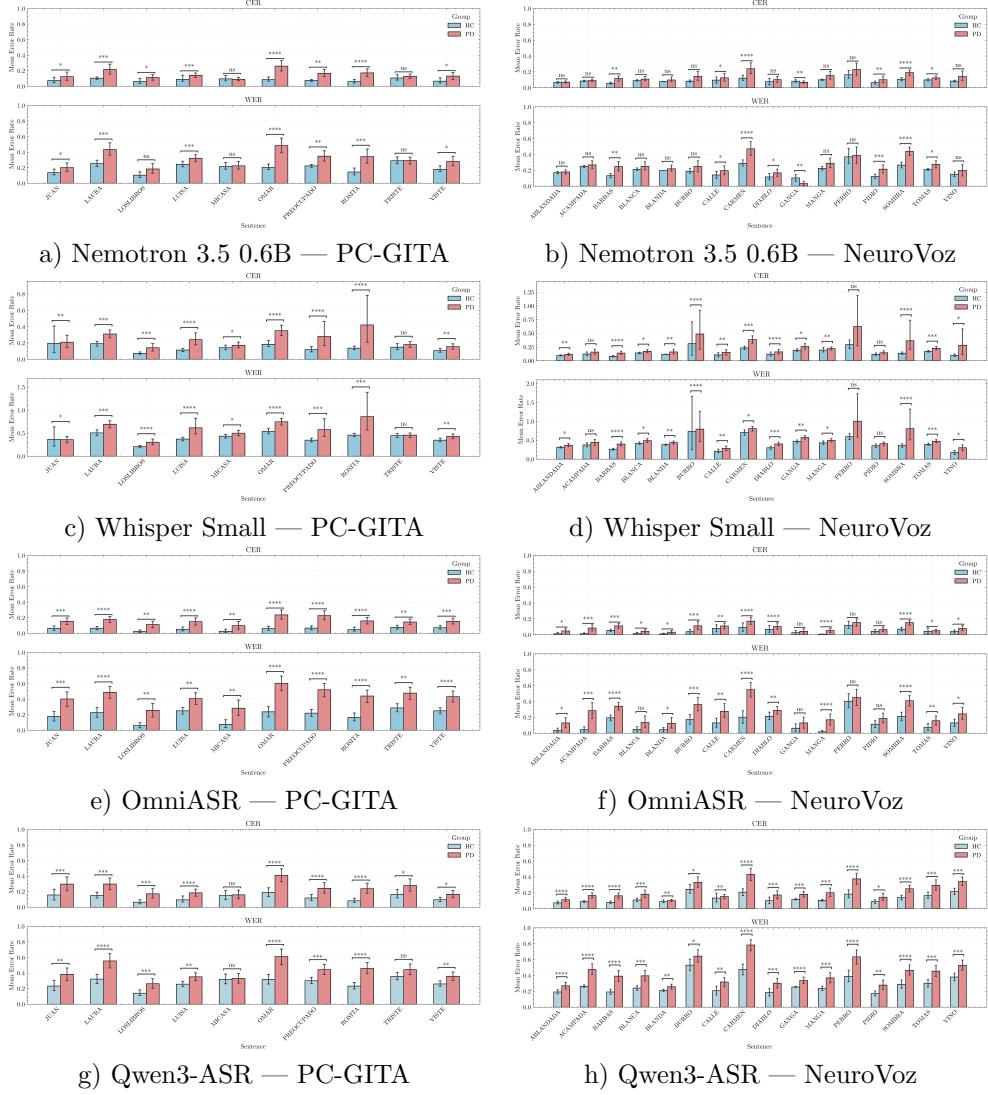


Fig. S2 Utterance-level ASR decoding errors (Word Error Rate, WER, and Character Error Rate, CER) across phonetically balanced sentences for the PC-GITA (Colombian Spanish, left column) and NeuroVoz (Castilian Spanish, right column) cohorts. Panels show Nemotron 3.5 0.6B (a,b), Whisper Small (c,d), OmniASR 300M (e,f), and Qwen3-ASR (g,h). For each sentence, mean error rates are compared between utterances from participants with Parkinson's disease (PD) and healthy controls (HC); error bars denote 95% confidence intervals and asterisks denote Mann-Whitney U significance (ns: not significant, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, ****: $p < 0.0001$).

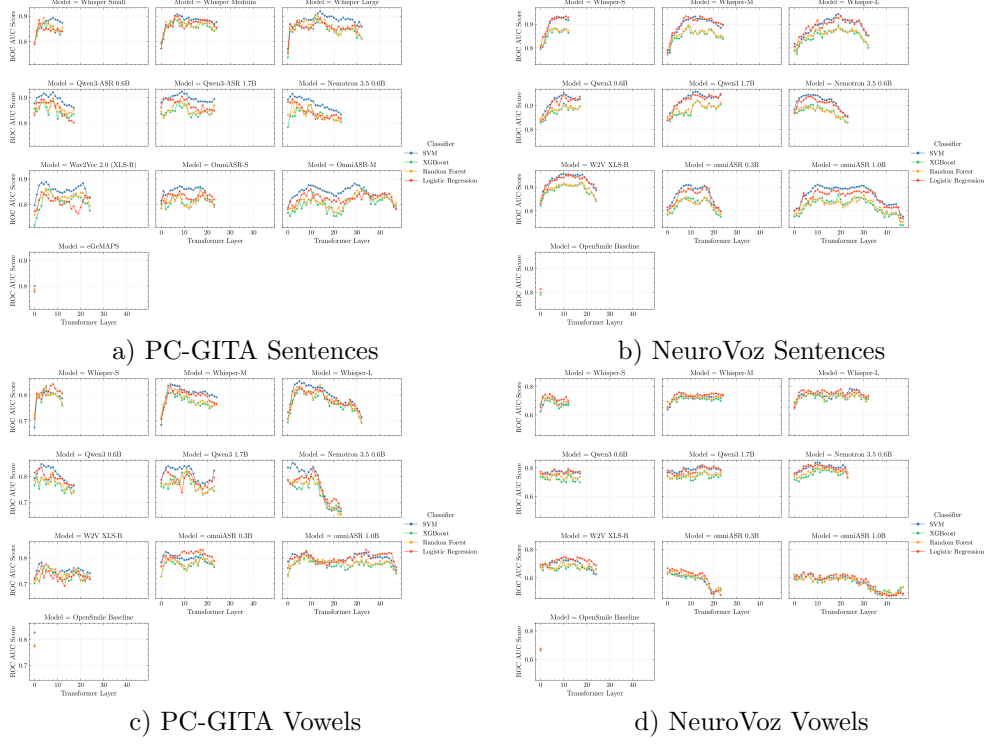


Fig. S5 Segment-level ROC-AUC as a function of extraction depth for each model and downstream classifier for PC-GITA sentence reading (a), NeuroVoz sentence reading (b), PC-GITA sustained vowels (c), and NeuroVoz sustained vowels (d). Initial embedding layers often perform less well than later representations, but the peak depth and the behavior of the deepest layers vary by model, classifier, task, and cohort.

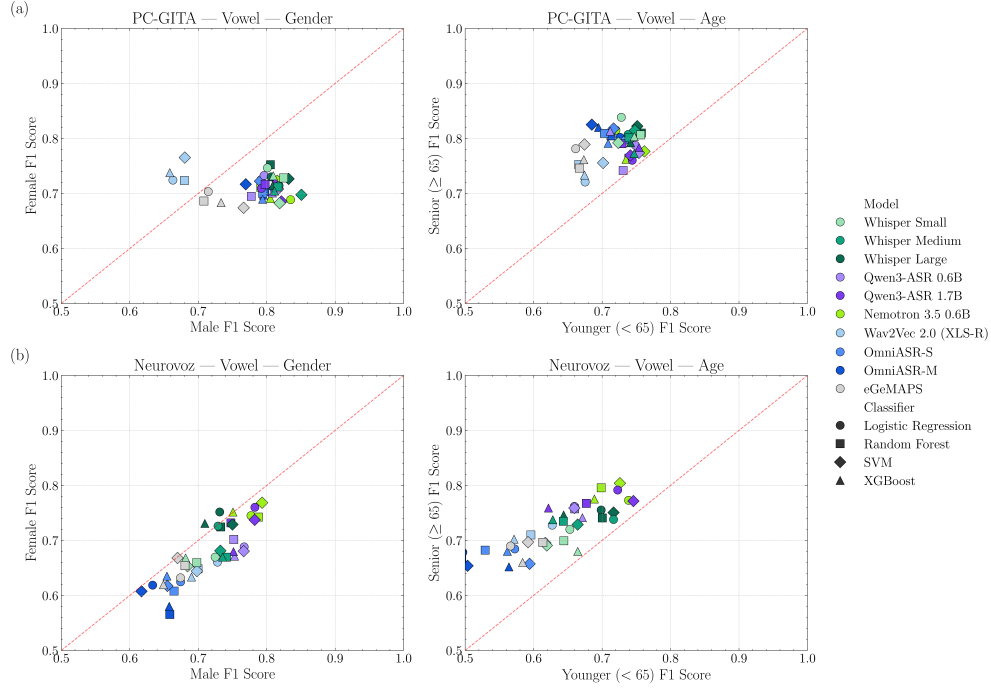


Fig. S6 Demographic parity scatterplots showing segment-level $F1$ -score across subgroups for the PC-GITA sustained vowel tasks (a) and NeuroVoz sustained vowel tasks (b). For each subplot, the left panel compares classification performance between genders (Male vs. Female), and the right panel compares performance across age strata (Younger (< 65) vs. Senior (≥ 65)). The red dashed diagonal line indicates demographic parity where classifier performance is identical between subgroups.

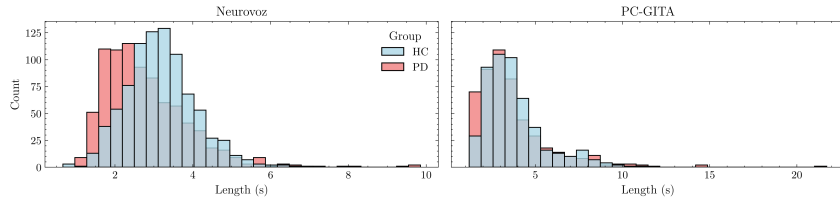


Fig. S7 Distributions of audio recording durations (in seconds) across Parkinson's disease (PD) participants and healthy controls (HC) for the PC-GITA and NeuroVoz sustained vowel and sentence reading tasks.

References

- [1] Mendes-Laureano, J., Gómez-García, J.A., Guerrero-López, A., Luque-Buzo, E., Arias-Londoño, J.D., Grandas-Pérez, F.J., Godino-Llorente, J.I.: Neurovoz: a castillian spanish corpus of parkinsonian speech. *Scientific Data* **11**(1), 1367 (2024) <https://doi.org/10.1038/s41597-024-04186-z>
- [2] Mendes-Laureano, J., Gómez-García, J.A., Guerrero-López, A., Luque-Buzo, E., Arias-Londoño, J.D., Grandas-Pérez, F.J., Godino-Llorente, J.I.: NeuroVoz: a Castillian Spanish corpus of parkinsonian speech. Zenodo. Restricted-access dataset released under CC BY-NC-ND 4.0 subject to a Data Usage Agreement (2024). <https://doi.org/10.5281/zenodo.10777656>
- [3] Orozco-Arroyave, J.R., Arias-Londoño, J.D., Vargas-Bonilla, J.F., González-Rátiva, M.C., Nöth, E.: New spanish speech corpus database for the analysis of people suffering from parkinson’s disease. In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC’14)*, pp. 342–347. European Language Resources Association (ELRA), Reykjavik, Iceland (2014)
- [4] Eyben, F., Wöllmer, M., Schuller, B.: Opensmile: The munich versatile and fast open-source audio feature extractor. In: *Proceedings of the 18th ACM International Conference on Multimedia. MM ’10*, pp. 1459–1462. Association for Computing Machinery, New York, NY, USA (2010). <https://doi.org/10.1145/1873951.1874246>
- [5] Eyben, F., Scherer, K.R., Schuller, B.W., Sundberg, J., Andre, E., Busso, C., Devillers, L.Y., Epps, J., Laukka, P., Narayanan, S.S., Truong, K.P.: The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing. *IEEE Transactions on Affective Computing* **7**(2), 190–202 (2016) <https://doi.org/10.1109/TAFFC.2015.2457417>